Estimatori în procesarea semnalelor

Conf. dr. habil. Eduard Rotenştein

1 Considerente asupra mărginii inferioare Rao-Cramer (CRLB)

După cum am studiat la Metoda verosimilităţii maxime, determinarea unei margini inferioare pentru dispersia unui estimator nedeplasat se dovedeşte a fi foarte importantă în practică. **În cel mai bun caz**, se poate stabili dacă un estimator nedeplasat este de dispersie minimă (*estimator MVU - Minimum variance unbiased estimator).* **În cel mai rău caz,** furnizează un reper în raport cu care putem compara performanţa oricărui alt estimator nedeplasat pentru parametrul estimat. Deşi există metode alternative de determinare a unei margini inferioare pentru dispersie (vezi McAulay, Hofstetter [1971], Kendall, Stuart [1979], Seidman [1970], Ziv, Zakai [1969]), marginea inferioară Rao-Cramer (*CRLB*) este cea mai uşor de obţinut. De asemenea, teoria ce conduce la stabilirea sa permite determinarea imediată a existenţei unui estimator ce o atinge. Chiar dacă nu există un astfel de estimator, se pot determina estimatori ce ating acea margine într-un sens aproximant.

1.1 Acurateţea estimatorului

Cum toate informaţiile privitoare la statistica analizată sunt încorporate în eşantionul observat şi în densitatea de repartiţie a caracteristicii, este evident că acurateţea estimării va depinde în mod direct de repartiţia urmată. Este, prin urmare, clar că nu putem vorbi de acurateţea estimării unui parametru dacă densitatea nu depinde de acesta sau depinde în mod slab de el.

Exemplul 1.1 ***Dependenţa densităţii de repartiţie de parametrul necunoscut.*** *Presupunem că, la o simplă observaţie, avem structura variabilei de selecţie dată de*

$$X_0 = A + W_0, \quad \textit{unde} \quad W_0 \sim \mathcal{N}\left(0, \sigma^2\right)$$

şi dorim să estimăm parametrul A. Desigur, este aşteptată o estimare cu atât mai bună cu cât dispersia σ^2 este mai mică. Într-adevăr, un estimator nedeplasat bun este $\hat{A} = X_0$, varianţa fiind σ^2, deci acurateţea estimatorului creşte odată cu descreşterea cantităţii σ^2. Un mod alternativ, exemplificativ, de a observa aceasta este considerarea a două densităţi de repartiţie, corespunzătoare la două dispersii $\sigma_1^2 < \sigma_2^2$:

$$f_i\left(x_0, A\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}\left(x_0 - A\right)^2\right), \quad i \in \{1, 2\} \tag{1}$$

Dacă reprezentăm cele două grafice pentru $x_0 = 3$ şi $\sigma_1 = 1/3$, atunci se observă pe reprezentare că valori $A > 4$ sunt puternic improbabile. Pentru a vedea aceasta,

$$\mathbb{P}\left(3 - \delta/2 \leq X_0 \leq 3 + \delta/2\right) = \int_{3-\delta/2}^{3+\delta/2} f_i\left(t, A\right) dt,$$

care, pentru δ suficient de mic, este $f_i\left(x_0 = 3, A\right)\delta$. Dar

$$f_1\left(x_0 = 3, A = 4\right)\delta = 0.10\delta, \quad \textit{în} \quad \textit{timp} \quad \textit{ce} \quad f_1\left(x_0 = 3, A = 3\right)\delta = 1.2\delta.$$

Valorile parametrului $A > 4$ pot fi deci eliminate din analiză.

Aplicând teoria verosimilităţii maxime, considerăm logaritmul funcţiei de verosimilitate:

$$\ln f\left(x_0, A\right) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\left(x_0 - A\right)^2, \quad \textit{pentru care}$$

$$\frac{\partial \ln f\left(x_0, A\right)}{\partial A} = \frac{1}{\sigma^2}\left(x_0 - A\right) \qquad \textit{şi} \qquad -\frac{\partial^2 \ln f\left(x_0, A\right)}{\partial A^2} = \frac{1}{\sigma^2} \tag{2}$$

Curbura graficului creşte pe măsură ce σ^2 descreşte. Cum ştiam deja că estimatorul $\hat{A} = X_0$ are dispersia σ^2, atunci, pentru acest exemplu,

$$D^2(\hat{A}) = \frac{1}{-\dfrac{\partial^2 \ln f\left(x_0, A\right)}{\partial A^2}} = \sigma^2.$$

Deşi în acest exemplu, derivata secundă nu depinde de valoarea empirică observată x_0, în general, va depinde. Din acest motiv, o măsură mai bună pentru curbura medie a graficului logaritmului funcţiei de verosimilitate este

$$-\mathbb{E}\left(\frac{\partial^2 \ln f\left(X_0, A\right)}{\partial A^2}\right). \tag{3}$$

Cu cât este mai mare cantitatea dată de (3), cu atât va fi mai mică dispersia estimatorului.

Revenind la aspectele teoretice prezentate la metoda verosimilităţii maxime, reamintim că, dacă densitatea de repartitie a caracteristicii studiate X este $f\left(x, \theta\right)$, atunci dispersia oricărui estimator nedeplasat $\hat{\theta}$ satisface relaţia

$$D^2(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left(\frac{\partial^2 \ln f\left(X, \theta\right)}{\partial \theta^2}\right)} = \frac{1}{I_1\left(\theta\right)} =: \frac{1}{I\left(\theta\right)} = CRLB(\hat{\theta}) \tag{4}$$

În plus, orice estimator nedeplasat care îşi atinge marginea (dispersiei) pentru toate valorile lui θ trebuie să satisfacă, ca şi rezultat de caracterizare, relaţia:

$$\frac{\partial \ln f\left(x, \theta\right)}{\partial \theta} = I\left(\theta\right)\left(g\left(x\right) - \theta\right), \tag{5}$$

pentru funcţiile convenabil alese g şi I. Acest estimator, care este un estimator *MVU este de fapt*

$$\hat{\theta} = g\left(X\right),$$

iar dispersia minimă atinsă este $1/I\left(\theta\right)$.

Exemplul 1.2 *CRLB pentru Exemplul 1.1. Din formulele (2) şi (4) obţinem că*

$$D^2(\hat{A}) \geq \sigma^2,$$

pentru toate valorile parametrului A. Prin urmare, nu poate exista niciun estimator nedeplasat a cărui dispersie să fie mai mică decât σ^2, nici măcar pentru o singură valoare a parametrului A.

$$\textit{Dar} \quad \hat{A} = X_0 \quad \textit{şi} \quad \textit{atunci} \quad D^2(\hat{A}) = \sigma^2.$$

Cum X_0 este nedeplasat şi atinge CRLB, va fi un estimator MVU. Din (2) şi (5) facem identificările:

$$\theta = A, \qquad I\left(\theta\right) = \frac{1}{\sigma^2} \qquad \textit{şi} \qquad g\left(X_0\right) = X_0.$$

Prin urmare, $\hat{A} = g\left(X_0\right) = X_0$, deci este estimator MVU. De asemenea,

$$D^2(\hat{A}) = \sigma^2 = \frac{1}{I\left(\theta\right)}$$

şi, în conformitate cu (4), avem

$$I\left(\theta\right) = -\mathbb{E}\left(\frac{\partial^2 \ln f\left(X, \theta\right)}{\partial \theta^2}\right).$$

Exemplul 1.3 ***Mărimea (amplitudinea) curentului continuu în prezenţa zgomotului alb Gaussian (WGN, White Gaussian Noise).*** *Generalizăm Exemplul 1.1 considerând cazul observaţiilor multiple date de variabilele de selecţie:*

$$X_i = A + W_i, \quad i \in \{0, 1, ..., n-1\}, \quad \textit{unde} \quad W_i \textit{ este WGN, de dispersie } \sigma^2, \ \forall i$$

Pentru determinarea CRLB pentru parametrul A, considerăm funcţia de verosimilitate

$$\begin{aligned} \mathcal{L}\left(V, A\right) = \mathcal{L}\left(X_0, ..., X_{n-1}, A\right) &= \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(X_i - A\right)^2\right) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}\left(X_i - A\right)^2\right), \end{aligned}$$

pentru care obţinem,

$$\frac{\partial \ln \mathcal{L}(V,A)}{\partial A} = \frac{\partial}{\partial A}\left(-\ln\left(\left(2\pi\sigma^2\right)^{n/2}\right) - \frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(X_i - A)^2\right) = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}(X_i - A) = \frac{n}{\sigma^2}\left(\bar{X} - A\right), \tag{6}$$

iar derivata secundă este constantă

$$\frac{\partial^2 \ln \mathcal{L}(V,A)}{\partial \theta^2} = -\frac{n}{\sigma^2}.$$

De asemenea, din (6) rezultă că, anulând derivata de oridinul întâi, găsim estimatorul $\hat{A} = \bar{X}$.

Formula (4) conduce la următoarea cantitate pentru CRLB:

$$D^2(\hat{A}) = D^2(\bar{X}) \geq \frac{\sigma^2}{n} = CRLB(\hat{A}), \tag{7}$$

iar media de selecţie este un estimator ce atinge acest prag minim, şi deci, prin urmare, este un estimator MVU. El este, deci, şi un estimator absolut corect.

Reamintim, de asemenea, că atunci când *CRLB* este atinsă, are loc egalitatea

$$D^2(\hat{\theta}) = \frac{1}{I(\theta)}, \quad \text{unde} \quad I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial \theta^2}\right).$$

Următorul exemplu arată că, condiţia de margine inferioară *CRLB* pentru dispersie nu este întotdeauna satisfăcută.

Exemplul 1.4 ***Estimarea fazei unui curent de tip sinusoidal într-un WGN.*** *Considerăm problema determinării unui estimator pentru parametrul fazei* ϕ*, al unui curent sinusoidal, aflată într-un WGN:*

$$X_i = A\cos(2\pi f_0 i + \phi) + W_i, \quad i \in \{0, 1, ..., n-1\}. \tag{8}$$

Amplitudinea A *şi frecvenţa* f_0 *se presupun a fi cunoscute. Densitatea de repartiţie a vectorului datelor este:*

$$\mathcal{L}(V,\phi) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}\left(X_i - A\cos(2\pi f_0 i + \phi)\right)^2\right).$$

Derivatele logaritmului funcţiei de verosimilitate sunt:

$$\begin{cases} \dfrac{\partial \ln \mathcal{L}(V,\phi)}{\partial \phi} &= -\dfrac{1}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}\left(X_i - A\cos(2\pi f_0 i + \phi)\right)A\sin(2\pi f_0 i + \phi) \\ &= -\dfrac{A}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}\left(X_i\sin(2\pi f_0 i + \phi) - \dfrac{A}{2}\sin(4\pi f_0 i + 2\phi)\right) \\ \dfrac{\partial^2 \ln \mathcal{L}(V,\phi)}{\partial \phi^2} &= -\dfrac{A}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}\left(X_i\cos(2\pi f_0 i + \phi) - A\cos(4\pi f_0 i + 2\phi)\right). \end{cases}$$

Având în vedere faptul că

$$\mathbb{E}(X_i) = \mathbb{E}(A\cos(2\pi f_0 i + \phi) + W_i) = A\cos(2\pi f_0 i + \phi) + \mathbb{E}(W_i) = A\cos(2\pi f_0 i + \phi),$$

şi relaţia $\cos(2x) = 2\cos^2(x) - 1$*, obţinem astfel următoarea formulă pentru informaţia Fisher corespunzătoare parametrului* ϕ *:*

$$\begin{aligned} -\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(V,\phi)}{\partial \phi^2}\right) &= \frac{A}{\sigma^2}\sum_{i=0}^{n-1}\left(A\cos^2(2\pi f_0 i + \phi) - A\cos(4\pi f_0 i + 2\phi)\right) \\ &= \frac{A^2}{\sigma^2}\sum_{i=0}^{n-1}\left(\frac{1}{2} + \frac{1}{2}\cos(4\pi f_0 i + 2\phi) - \cos(4\pi f_0 i + 2\phi)\right) \simeq \frac{nA^2}{2\sigma^2}, \end{aligned}$$

deoarece

$$\frac{1}{n}\sum_{i=0}^{n-1}\cos(4\pi f_0 i + 2\phi) \simeq 0 \quad \textit{pentru} \quad f_0 \textit{ neapropiat de } 0 \textit{ sau } 1/2.$$

Prin urmare,

$$D^2(\hat{\phi}) > \frac{2\sigma^2}{nA^2}.$$

În acest exemplu, condiţia pentru existenţa marginii inferioare atinse nu este satisfăcută. În consecinţă, ***nu există un estimator pentru fază care să fie nedeplasat şi care să atingă CRLB.***

Este posibil însă ca să existe un estimator de dispersie minimă, MVU. Nu putem, pentru moment, să stabilim dacă un estimator MVU pentru parametru există sau nu, iar dacă el există, cum să îl determinăm. Teoria Statisticilor suficiente ne va permite să găsim răspunsuri la aceste întrebări.

Aşa cum am văzut, un estimator nedeplasat care atinge CRLB este un estimator eficient. Un estimator MVU nu este, neapărat, şi eficient. Prin urmare, dacă marginea CRLB este atinsă, dispersia estimatorului este inversa informaţiei Fisher. Intuitiv, cu cât avem la dispoziţie mai multe informaţii, cu atât marginea inferioară a dispersiei estimatorului este mai mică. Pentru observaţii **ne-independente**, este de aşteptat ca informaţia Fisher să fie mai mică decât $nI_1(\theta)$ .Pentru observaţii **complet dependente**,

$$I_n(\theta) = I_1(\theta),$$

adică observaţii suplimentare nu vor aduce informaţii noi, ceea ce va face ca *CRLB s*ă nu descrească odată cu creşterea numărului de date noi observate.

Exemplul anterior poate fi extins la situaţia generală a determinării marginii inferioare a estimatorului $\hat{\theta}$ pentru semnale electrice de tipul următor, în care nu se dă forma explicită a dependenţei semnalului de parametrul θ :

$$X_i = s_i(\theta) + W_i, \quad i \in \{0, 1, ..., n-1\}.$$

Funcţia de verosimilitate este, deoarece $X_i \sim \mathcal{N}\left(s_i(\theta), \sigma^2\right)$,

$$\mathcal{L}(V, \phi) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}\left(X_i - s_i(\theta)\right)^2\right),$$

iar diferenţialele logaritmului său sunt:

$$\begin{cases} \dfrac{\partial \ln \mathcal{L}(V,\phi)}{\partial \phi} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}\left(X_i - s_i(\theta)\right)\frac{\partial s_i(\theta)}{\partial \theta} \\ \dfrac{\partial^2 \ln \mathcal{L}(V,\phi)}{\partial \phi^2} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}\left(\left(X_i - s_i(\theta)\right)\frac{\partial^2 s_i(\theta)}{\partial \theta^2} - \left(\frac{\partial s_i(\theta)}{\partial \theta}\right)^2\right). \end{cases}$$

Aplicăm media şi obţinem, deoarece $\mathbb{E}\left(X_i - s_i(\theta)\right) = \mathbb{E}(W_i) = 0$, iar $\partial^2 s_i(\theta)/\partial\theta^2$ este deterministă,

$$\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(V,\phi)}{\partial \phi^2}\right) = -\frac{1}{\sigma^2}\sum_{i=0}^{n-1}\left(\frac{\partial s_i(\theta)}{\partial \theta}\right)^2.$$

În consecinţă,

$$D^2(\hat{\theta}) \geq \frac{\sigma^2}{\displaystyle\sum_{i=0}^{n-1}\left(\frac{\partial s_i(\theta)}{\partial \theta}\right)^2} = \frac{1}{I(\theta)}. \tag{9}$$

În particular, dacă semnalul curentului sinusoidal este reprezentată de (8), cu s_i depinzând de frecvenţa f_0 în modul următor:

$$s_i(f_0) := A\cos(2\pi f_0 i + \phi), \quad f_0 \in (0, 1/2),$$

amplitudinea A şi faza ϕ fiind cunoscute, atunci estimatorul frecvenţei curentului are marginea *CRLB*:

$$D^2(\hat{f}_0) \geq \frac{\sigma^2}{A^2\displaystyle\sum_{i=0}^{n-1}\left(2\pi i \sin(2\pi f_0 i + \phi)\right)^2} = CRLB(\hat{f}_0).$$

Menţionăm că, dacă

$$f_0 \to 0, \quad \text{atunci} \quad CRLB(\hat{f}_0) \longrightarrow +\infty.$$

Intuitiv, aceasta se datorează faptului că, pentru o frecvenţă f_0 suficient de mică (neglijabilă), o modificare a sa nu va altera semnalul într-o manieră semnificativă.

1.2 Transformarea parametrilor

De multe ori, în practică, se întâmplă ca parametrii pe care dorim să îi estimăm să depindă, sub forma unei funcţii de alţi parametri. De exemplu, putem fi interesaţi nu de estimarea nivelului curentului A, ci de puterea semnalului, A^2. În general, dacă dorim să estimăm parametrul $\alpha = g(\theta)$, atunci *CRLB este:*

$$D^2(\hat{\alpha}) \geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{-\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial \theta^2}\right)}. \tag{10}$$

Particularizând, în exemplul anterior, pentru $\alpha = g(A) = A^2$,

$$D^2\left(\widehat{A^2}\right) \geq \frac{(2A)^2}{n/\sigma^2} = \frac{4A^2\sigma^2}{n} = CRLB\left(\widehat{A^2}\right). \tag{11}$$

În Exemplul 1.3, media de selecţie era un estimator eficient pentru A. Am putea presupune că $\bar{X}^2$ este, de asemenea, un estimator eficient şi pentru A^2. Pentru a evita această eroare de raţionament arătăm, pentru început, că $\bar{X}^2$ nu este nici măcar un estimator nedeplasat. Deoarece

$$\bar{X} \sim \mathcal{N}\left(A, \sigma^2/n\right),$$

avem:

$$\mathbb{E}(\bar{X}^2) = \mathbb{E}^2(\bar{X}) + D^2(\bar{X}) = A^2 + \frac{\sigma^2}{n} \neq A^2. \tag{12}$$

Putem deci afirma imediat că **eficienţa unui estimator este distrusă de o transformare neliniară a sa**. **Eficienţa se păstrează însă în cazul transformărilor liniare**. Pentru a verifica această ultimă afirmaţie, presupunem că există un estimator eficient $\hat{\theta}$, pentru parametrul θ. Ne propunem acum să estimăm noul parametru

$$g(\theta) = a\theta + b, \quad \text{pentru} \quad a, b \in \mathbb{R}.$$

Alegem drept estimator statistica

$$\widehat{g(\theta)} := g(\hat{\theta}) = a\hat{\theta} + b.$$

Avem astfel:

$$\begin{cases} \mathbb{E}\left(\widehat{g(\theta)}\right) & = & \mathbb{E}(a\hat{\theta} + b) = a\mathbb{E}(\hat{\theta}) + b = a\theta + b = g(\theta), \quad \text{deci } \widehat{g(\theta)} \text{ este nedeplasat;} \\ D^2\left(\widehat{g(\theta)}\right) & \overset{(10)}{\geq} & \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)} = \left(\frac{\partial g}{\partial \theta}\right)^2 D^2(\hat{\theta}) = a^2 D^2(\hat{\theta}) = D^2(a\hat{\theta} + b) = D^2\left(\widehat{g(\theta)}\right), \end{cases}$$

ceea ce implică faptul că are loc egalitate în ultima relaţie, adică *CRLB* este atinsă.

Deşi eficienţa se conservă în cazul transformărilor liniare ale parametrilor, ea este păstrată "aproximativ" şi în cazul transformărilor neliniare dacă eşantionul considerat este suficient de mare. Aceasta are semnificaţie din punct de vedere practic deoarece se doreşte frecvent estimarea unor funcţii generale dependente de parametrul de bază. Pentru a vizualiza afirmaţia anterioară revenim la exemplul estimării lui A^2 cu ajutorul statisticii $\bar{X}^2$. Cu toate că aceasta este deplasată, din (12) se poate observa că, asimptotic, este de fapt un estimator nedeplasat (atunci când $n \to +\infty$). El este deci doar corect, nu şi absolut corect. În plus, cum

$$\bar{X} \sim \mathcal{N}\left(A, \sigma^2/n\right),$$

putem evalua dispersia

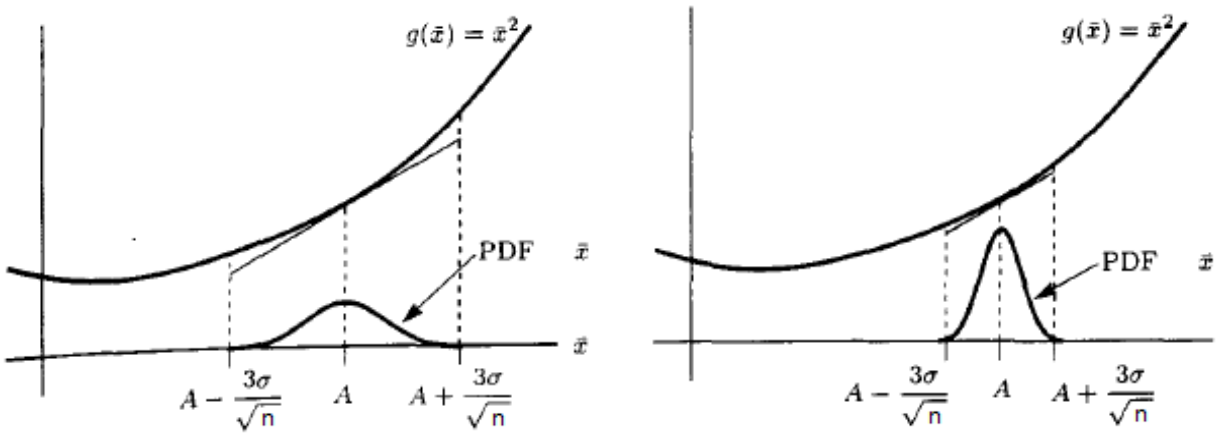
$$D^2(\bar{X}^2) = \mathbb{E}\left(\bar{X}^4\right) - \mathbb{E}^2\left(\bar{X}^2\right) = \frac{4A^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}. \tag{13}$$

Pentru ultima egalitate, am folosit faptul că, dacă $\xi \sim \mathcal{N}\left(\mu, \sigma^2\right)$, atunci

$$\mathbb{E}\left(\xi^2\right) = \mu^2 + \sigma^2, \qquad \mathbb{E}\left(\xi^4\right) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4,$$

ceea ce conduce la

$$D^2(\xi^2) = \mathbb{E}\left(\xi^4\right) - \mathbb{E}^2\left(\xi^2\right) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - (\mu^4 + \sigma^4 + 2\mu^2\sigma^2) = 4\mu^2\sigma^2 + 2\sigma^4.$$



(a) n mic Figura 1. (b) n mare

Se observă că al doilea termen din (13) converge la zero mai repede decât primul, ceea ce face ca, asimptotic, să regăsim *CRLB* din inegalitatea (11). Pentru un volum de selecţie mare, media de selecţie este concentrată în vecinătatea mediei teoretice A, iar pe acest interval mic transformarea neliniară este aproximativ liniară. De fapt, dacă aproximăm, prin liniarizare, funcţia g în vecinătatea lui A, obţinem:

$$g(\bar{x}) \simeq g(A) + g'(A)(\bar{x} - A), \quad \text{iar}$$

$$\mathbb{E}(g(\bar{X})) = g(A) = A^2, \quad \text{deoarece} \quad \mathbb{E}\left(\bar{X} - A\right) = \mathbb{E}\left(\bar{X}\right) - A = 0,$$

estimatorul fiind deci nedeplasat. De asemenea,

$$D^2(g(\bar{X})) = \left(g'(A)\right)^2 D^2(\bar{X}) = \frac{(2A)^2 \sigma^2}{n} = \frac{4A^2\sigma^2}{n},$$

adică estimatorul atinge, într-adevăr, **asimptotic**, *CRLB*.

1.3 Cazul parametrilor multipli (vectoriali)

Extindem analiza la cazul în care caracteristica studiată este dependentă de mai mulţi parametri, vectorul acestora fiind notat cu $\theta = (\theta_1, \theta_2, ..., \theta_p)^t$. Presupunem, de asemenea, că estimatorul $\hat{\theta}$ este nedeplasat. Marginea inferioară a dispersiei va fi un vector *CRLB* ce impune o limitare inferioară a fiecărei componente. Vom obţine

$$\left|\begin{array}{l} D^2(\hat{\theta}_i) \geq (I^{-1}(\theta))_{ii}, \quad \text{unde} \quad I(\theta) \in \mathcal{M}_{p\times p}(\mathbb{R}) \text{ este matricea informaţiei Fisher:} \\ \\ (I(\theta))_{ij} = -\mathbb{E}\left(\dfrac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial\theta_i \partial\theta_j}\right), \quad \text{pentru } i,j \in \{1,2,...,p\}. \end{array}\right. \tag{14}$$

Exemplul 1.5 ***Analiza curentului continuu bi-parametrizat în WGN.*** *Revenim la Exemplul 1.3 şi presupunem că, atât A, cât şi σ^2 sunt necunoscute. Vectorul parametrilor este deci*

$$\theta = (A, \sigma^2)^t, \quad p = 2.$$

Matricea informaţiei Fisher asociate are forma:

$$I(\theta) = \begin{pmatrix} -\mathbb{E}\left(\dfrac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A^2}\right) & -\mathbb{E}\left(\dfrac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A \partial\sigma^2}\right) \\ \\ -\mathbb{E}\left(\dfrac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial\sigma^2 \partial A}\right) & -\mathbb{E}\left(\dfrac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial(\sigma^2)^2}\right) \end{pmatrix}$$

Logaritmul funcţiei de verosimilitate şi derivatele parţiale ale acestuia sunt:

$$\ln \mathcal{L}(V,\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(X_i - A)^2$$

$$\frac{\partial \ln \mathcal{L}(V,\theta)}{\partial A} = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}(X_i - A), \qquad \frac{\partial \ln \mathcal{L}(V,\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=0}^{n-1}(X_i - A)^2, \quad \frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A^2} = -\frac{n}{\sigma^2},$$

$$\frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=0}^{n-1}(X_i - A), \quad \frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=0}^{n-1}(X_i - A)^2.$$

Matricea informaţiei Fisher devine:

$$I(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

Pentru ultimul element din matricea anterioară am folosit evaluarea:

$$\begin{aligned} -\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial (\sigma^2)^2}\right) &= -\mathbb{E}\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=0}^{n-1}(X_i - A)^2\right) = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=0}^{n-1}\mathbb{E}(X_i - A)^2 \\ &= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=0}^{n-1}\mathbb{E}(W_i^2) = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=0}^{n-1}\left(D^2(W_i) + \mathbb{E}^2(W_i)\right) \\ &= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=0}^{n-1}(\sigma^2 + 0) = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^4}n = \frac{n}{2\sigma^4}. \end{aligned}$$

Din forma anterioară determinăm marginile

$$D^2(\hat{A}) \geq \frac{\sigma^2}{n} \qquad şi \qquad D^2(\widehat{\sigma^2}) \geq \frac{2\sigma^4}{n}.$$

Observaţia 1.1 *Chiar dacă, pentru acest exemplu, matricea $I(\theta)$ este de tip diagonal, în cazul general ea nu are această proprietate. Remarcăm, de asemenea, că CRLB pentru estimatorul $\hat{A}$ este acelaşi ca şi cazul în care σ^2 era cunosut. Acest lucru se întâmplă doar datorită formei diagonale a matricei $I(\theta)$. În absenţa acestei proprietăţi, nici identificarea anterioară nu mai este posibilă, după cum vedem în exemplul următor.*

Exemplul 1.6 ***Aproximarea dreptei de regresie (line fitting).*** *Considerăm observaţiile date de următoarele formule de dependenţă, ce vor conduce la estimarea dreptei de regresie:*

$$X_i = A + Bi + W_i, \quad unde \quad W_i \text{ sunt WGN independente}, \quad i \in \{0, 1, ..., n-1\}.$$

Ne propunem să determinăm CRLB pentru parametrii A (intercept) şi B (slope), vectorul parametrilor fiind $\theta = (A, B)^t$. Funcţia de verosimilitate şi derivatele parţiale ale logaritmului său sunt, deoarece $X_i \sim \mathcal{N}(A + Bi, \sigma^2)$:

$$\mathcal{L}(V,\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(X_i - A - Bi)^2\right)$$

$$\frac{\partial \ln \mathcal{L}(V,\theta)}{\partial A} = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}(X_i - A - Bi), \quad \frac{\partial \ln \mathcal{L}(V,\theta)}{\partial B} = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}(X_i - A - Bi)\,i, \quad \frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A^2} = -\frac{n}{\sigma^2},$$

$$\frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial A \partial B} = -\frac{1}{\sigma^2}\sum_{i=0}^{n-1} i, \qquad \frac{\partial^2 \ln \mathcal{L}(V,\theta)}{\partial B^2} = -\frac{1}{\sigma^2}\sum_{i=0}^{n-1} i^2.$$

Matricea informaţiei Fisher are următoarea forma imediată, deoarece derivatele secunde nu depind de variabilele X_i :

$$I(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} n & \sum_{i=0}^{n-1} i \\ \sum_{i=0}^{n-1} i & \sum_{i=0}^{n-1} i^2 \end{pmatrix} = \frac{1}{\sigma^2}\begin{pmatrix} n & \frac{n(n-1)}{2} \\ \frac{n(n-1)}{2} & \frac{n(n-1)(2n-1)}{6} \end{pmatrix}.$$

Inversa acestei matrici este:

$$I^{-1}(\theta)=\sigma^2\begin{pmatrix}\frac{2\,(2n-1)}{n\,(n+1)} & -\frac{6}{n\,(n+1)}\\ -\frac{6}{n\,(n+1)} & \frac{12}{n(n^2-1)}\end{pmatrix}.$$

Formula (14) arată că CRLB sunt date de inegalităţile

$$D^2(\hat{A})\geq\frac{2\,(2n-1)\,\sigma^2}{n\,(n+1)}=CRLB(\hat{A})\qquad şi\qquad D^2(\hat{B})\geq\frac{12\sigma^2}{n(n^2-1)}=CRLB(\hat{B}).$$

Remarcăm faptul că CRLB pentru estimatorul $\hat{A}$ creşte faţă de situaţia în care parametrul B era cunoscut, deoarece, pentru $n\geq 2$,

$$D^2(\hat{A})\geq\frac{2\,(2n-1)}{n+1}\cdot\frac{\sigma^2}{n}\geq 1\cdot\frac{\sigma^2}{n}=-\frac{1}{\mathbb{E}\left(\frac{\partial^2\ln\mathcal{L}\,(V,A)}{\partial A^2}\right)},$$

ultimul termen identificându-se chiar cu CRLB obţinută pentru $\hat{A}$, când B era cunoscut. Anticipăm astfel un rezultat general valabil, şi anume faptul CRLB întotdeauna creşte odată cu creşterea numărului de parametri estimaţi.

În altă ordine de idei,

$$\frac{CRLB(\hat{A})}{CRLB(\hat{B})}=\frac{\frac{2\,(2n-1)\,\sigma^2}{n\,(n+1)}}{\frac{12\sigma^2}{n(n^2-1)}}=\frac{(2n-1)\,(n-1)}{6}\geq 1,\quad \textit{pentru } n\geq 3.$$

Prin urmare, parametrul B este mai uşor de estimat, marginea sa inferioară CRLB având o descreştere de ordinul lui $1/n^3$, în comparaţie cu dependenţa de ordinul $1/n$ a marginii CRLB a celuilalt parametru. ***Această diferenţa de magnitudine indică faptul că observaţia X_i este mai senzitivă la modificări ale parametrului B faţă de modificări ale parametrului A.*** *Observăm, de asemenea, că*

$$\Delta X_i\simeq\frac{\partial X_i}{\partial A}\Delta A=\Delta A\qquad şi\qquad \Delta X_i\simeq\frac{\partial X_i}{\partial B}\Delta B=i\Delta B,$$

adică modificările lui B se resimt multiplicate în evoluţia observaţiilor (aleatoare) X_i.

Prezentăm acum varianta vectorială a Teoremei Rao-Cramer pentru marginea inferioară.

Teorema 1 *Presupunem că densitatea de repartiţie a vectorului aleator de selecţie $\mathcal{L}\,(V,\theta)$ satisface următoarele condiţii de regularitate:*

$$\mathbb{E}\left(\frac{\partial\ln\mathcal{L}\,(V,\theta)}{\partial\theta}\right)=0,\quad \textit{pentru toţi parametrii } \theta.$$

Atunci, matricea de covarianţă $C_{\hat{\theta}}$ a oricărui estimator nedeplasat $\hat{\theta}$ satisface inegalitatea

$$C_{\hat{\theta}}-I^{-1}(\hat{\theta})\geq\mathbf{0},\tag{15}$$

inegalitatea fiind înţeleasă în sensul că matricea din membrul stâng este pozitiv semi-definită. Informaţia Fisher este matricea

$$(I(\theta))_{ij}=-\mathbb{E}\left(\frac{\partial^2\ln\mathcal{L}\,(V,\theta)}{\partial\theta_i\partial\theta_j}\right).$$

În plus, un estimator nedeplasat atinge marginea inferioară în sensul că

$$C_{\hat{\theta}}=I^{-1}(\hat{\theta})$$

dacă şi numai dacă

$$\frac{\partial\ln\mathcal{L}\,(V,\theta)}{\partial\theta}=I(\theta)\,(g(V)-\theta)\,,\tag{16}$$

pentru o funcţie $g:\mathbb{R}^n\to\mathbb{R}^p$ şi o matrice $I\in\mathcal{M}_{p\times p}$.

Estimatorul, care este un estimator de dispersie minimă (MVU) este:

$$\hat{\theta}=g(V),$$

iar matricea sa de covarianţă este $I^{-1}(\theta)$.

Demonstraţie. Considerăm vectorul parametrilor $\alpha = g(\theta)$, densitatea de repartiţie a caracteristicii fiind dependentă de vectorul θ al parametrilor. Fie estimatorii nedeplasaţi astfel încât

$$\mathbb{E}(\hat{\alpha}_i) = \alpha_i = (g(\theta))_i, \quad i \in \{1, 2, ..., r\}.$$

Condiţiile de regularitate conduc la

$$\int_{\mathbb{R}^n} (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta_i} \mathcal{L}(v,\theta)\, dv = \frac{\partial (g(\theta))_i}{\partial \theta_i}, \quad i \in \{1, 2, ..., p\}. \tag{17}$$

Pentru $j \neq i$,

$$\begin{aligned} &\int_{\mathbb{R}^n} (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta_j} \mathcal{L}(v,\theta)\, dv = \int_{\mathbb{R}^n} (\hat{\alpha}_i - \alpha_i) \frac{\partial \mathcal{L}(v,\theta)}{\partial \theta_j} dv \\ &= \frac{\partial}{\partial \theta_j} \int_{\mathbb{R}^n} \hat{\alpha}_i \mathcal{L}(v,\theta)\, dv - \alpha_i \underbrace{\mathbb{E}\left(\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta_j} \right)}_{=0} = \frac{\partial \alpha_i}{\partial \theta_j} = \frac{\partial (g(\theta))_i}{\partial \theta_j} \end{aligned} \tag{18a}$$

Scriind (17) şi (18a) sub formă matriceală obţinem

$$\int_{\mathbb{R}^n} (\hat{\alpha} - \alpha) \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathcal{L}(v,\theta)\, dv = \frac{\partial g(\theta)}{\partial \theta}.$$

Considerăm acum vectorii, posibili arbitrari, $\mathbf{a} \in \mathcal{M}_{r\times 1}$ şi $\mathbf{b} \in \mathcal{M}_{p\times 1}$. Multiplicând în poziţiile convenabile (la stânga / la dreapta) cu vectorii $\mathbf{a}^t$ şi $\mathbf{b}$ şi ţinând cont de faptul că $\mathcal{L}(v,\theta)$ este un scalar, găsim:

$$\int_{\mathbb{R}^n} \underbrace{\mathbf{a}^t(\hat{\alpha} - \alpha)} \cdot \underbrace{\frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b}} \cdot \underbrace{\mathcal{L}(v,\theta)}\, dv = \mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} \mathbf{b}. \tag{19}$$

Notăm

$$w(v) = \mathcal{L}(v,\theta), \qquad g(v) = \mathbf{a}^t(\hat{\alpha} - \alpha), \qquad h(v) = \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b}.$$

Ridicăm la pătrat (19) şi, aplicând inegalitatea Cauchy Schwarz, obţinem că:

$$\begin{aligned} \left(\mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} \mathbf{b} \right)^2 &\leq \int_{\mathbb{R}^n} \left(\sqrt{\mathcal{L}(v,\theta)} \mathbf{a}^t(\hat{\alpha} - \alpha) \right) \cdot \left(\sqrt{\mathcal{L}(v,\theta)} \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b} \right) dv \\ &\leq \int_{\mathbb{R}^n} \mathcal{L}(v,\theta)\, \mathbf{a}^t(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^t \mathbf{a} dv \cdot \int_{\mathbb{R}^n}^t \mathcal{L}(v,\theta)\, \mathbf{b}^t \frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta} \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b} dv \\ &= \mathbf{a}^t C_{\hat{\alpha}} \mathbf{a} \cdot \mathbf{b}^t I(\theta)\, \mathbf{b}, \end{aligned} \tag{20}$$

deoarece, similar cazului scalar,

$$\mathbb{E}\left(\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta_i} \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta_j} \right) = -\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(v,\theta)}{\partial \theta_i \partial \theta_j} \right) = (I(\theta))_{ij}.$$

Cum $\mathbf{b}$ este arbitrar ales, alegem pentru el valoarea

$$\mathbf{b} := I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \mathbf{a} \qquad \text{deci} \qquad \mathbf{b}^t = \mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} \left(I^{-1}(\theta) \right)^t = \mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta).$$

Obţinem din (20),

$$\begin{aligned} \left(\mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \mathbf{a} \right)^2 &\leq \mathbf{a}^t C_{\hat{\alpha}} \mathbf{a} \cdot \left(\mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) I(\theta)\, I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \mathbf{a} \right) \\ &= \mathbf{a}^t C_{\hat{\alpha}} \mathbf{a} \cdot \left(\mathbf{a}^t \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \mathbf{a} \right). \end{aligned}$$

Deoarece $I(\theta)$ este pozitiv definită, aceeaşi proprietate o are şi $I^{-1}(\theta)$, iar matricea

$$\frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta}$$

este, cel puţin, pozitiv semi-definită. Prin urmare, termenul din parantezele anterioare este ne-negativ şi deducem că

$$\mathbf{a}^t \left(C_{\hat{\alpha}} - \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \right) \mathbf{a} \geq 0.$$

Reamintim că şi $\mathbf{a}$ este ales arbitrar, ceea ce conduce la

$$C_{\hat{\alpha}} - \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \geq 0.$$

Dacă

$$\alpha = g(\theta) = \theta, \quad \text{atunci} \quad \frac{\partial g(\theta)}{\partial \theta} = I,$$

iar inegalitatea (15) este satisfăcută. Condiţia pentru egalitate în inegalitatea Cauchy Schwarz este

$$g(v) = c \cdot h(v) \quad \Longleftrightarrow \quad \mathbf{a}^t(\hat{\alpha} - \alpha) = c \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b},$$

unde c este o constantă independentă de v. Această condiţie devine:

$$\mathbf{a}^t(\hat{\alpha} - \alpha) = c \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \mathbf{b} = c \frac{\partial \ln \mathcal{L}(v,\theta)^t}{\partial \theta} \underbrace{I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \mathbf{a}}_{=\mathbf{b}}.$$

Cum vectorul $\mathbf{a}$ este arbitrar ales,

$$\frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta} = \frac{1}{c}(\hat{\alpha} - \alpha).$$

Dacă considerăm $\alpha = g(\theta) = \theta$, atunci

$$\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta} = \frac{1}{c} I(\theta)(\hat{\theta} - \theta).$$

Cum c poate depinde de vectorul parametrilor θ,

$$\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta_i} = \sum_{k=1}^{p} \frac{(I(\theta))_{ik}}{c(\theta)} (\hat{\theta}_k - \theta_k), \quad \text{de unde, diferenţiind încă o dată,}$$

$$\frac{\partial^2 \ln \mathcal{L}(v,\theta)}{\partial \theta_i \partial \theta_j} = \sum_{k=1}^{p} \left(\frac{(I(\theta))_{ik}}{c(\theta)} (-\delta_{kj}) + \frac{\partial \left(\frac{(I(\theta))_{ik}}{c(\theta)} \right)}{\partial \theta_j} (\hat{\theta}_k - \theta_k) \right).$$

În final,

$$(I(\theta))_{ij} = -\mathbb{E}\left(\frac{\partial^2 \ln \mathcal{L}(v,\theta)}{\partial \theta_i \partial \theta_j} \right) = \frac{(I(\theta))_{ij}}{c(\theta)}, \quad \text{deoarece} \quad \mathbb{E}(\hat{\theta}_k) = \theta_k.$$

Evident,

$$c(\theta) = 1,$$

iar condiţia pentru egalitate rezultă imediat. Demonstraţia este, în acest moment, încheiată.

Faptul că relaţia (14) rezultă din (15) se justifică remarcând că pentru o matrice pozitiv semi-definită elementele diagonale sunt ne-negative. Prin urmare,

$$(C_{\hat{\theta}} - I^{-1}(\theta))_{ii} \geq 0$$

şi, în consecinţă,

$$D^2(\hat{\theta}_i) \geq (C_{\hat{\theta}})_{ii} \geq (I^{-1}(\theta))_{ii}, \quad \text{pentru orice } i. \tag{21}$$

Când are loc egalitatea (adică $C_{\hat{\theta}} = I^{-1}(\theta)$), atunci şi (21) are loc cu egalitate. Estimatorul

$$\hat{\theta} = g(V)$$

este eficient şi, prin urmare, este un estimator *MVU*. Un caz privind egalitatea am obţinut in Exemplul 1.6. Am determinat

$$\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta} = \begin{pmatrix} \dfrac{\partial \ln \mathcal{L}(v,\theta)}{\partial A} \\ \\ \dfrac{\partial \ln \mathcal{L}(v,\theta)}{\partial B} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}(X_i - A - Bi) \\ \\ \dfrac{1}{\sigma^2}\displaystyle\sum_{i=0}^{n-1}(X_i - A - Bi)\, i \end{pmatrix}. \tag{22}$$

Putem scrie (22) sub caracterizarea echivalentă (se verifică prin calcul direct structura următoare)

$$\frac{\partial \ln \mathcal{L}(v,\theta)}{\partial \theta} = \begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{n(n-1)}{2\sigma^2} \\ \\ \dfrac{n(n-1)}{2\sigma^2} & \dfrac{n(n-1)(2n-1)}{6\sigma^2} \end{pmatrix} \begin{pmatrix} \hat{A} - A \\ \\ \hat{B} - B \end{pmatrix}, \tag{23}$$

unde

$$\hat{A} = \frac{2(2n-1)}{n(n+1)}\sum_{i=0}^{n-1} X_i - \frac{6}{n(n+1)}\sum_{i=0}^{n-1} iX_i \qquad \text{şi} \qquad \hat{B} = -\frac{6}{n(n+1)}\sum_{i=0}^{n-1} X_i + \frac{12}{n(n^2-1)}\sum_{i=0}^{n-1} iX_i.$$

Condiţiile pentru egalitate sunt satisfăcute, iar estimatorul

$$\hat{\theta} = (\hat{A}, \hat{B})^t$$

este eficient şi, prin urmare, este un estimator *MVU*. Matricea din (23) este inversa matricei de covarianţă.

Dacă condiţia de egalitate are loc, atunci, întotdeauna, estimatorul $\hat{\theta}$ este nedeplasat. Într-adevăr, condiţia de regularitate, aplicată reprezentării (16) implică

$$\mathbb{E}(g(V)) = \mathbb{E}(\hat{\theta}) = \theta.$$

Pentru determinarea estimatorilor *MVU* pentru un parametru vectorial, Teorema *CRLB* este un instrument recomandat. În general, metoda este folosită în aplicaţii practice, pentru modele liniare, Exemplul 1.6 al estimării dreptei de regresie (*line fitting*) fiind doar un caz particular.

1.4 Transformarea parametrilor vectoriali

Dorim să estimăm parametrul vectorial $\alpha = g(\theta)$, g fiind o funcţie $r-$dimensională, $g : \mathbb{R}^p \to \mathbb{R}^r$. După cum am demonstrat în Teorema 1,

$$C_{\hat{\alpha}} - \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} \geq 0. \tag{24}$$

Prin $\partial g(\theta)/\partial \theta$ am înţeles, desigur, matricea Jacobiană $r \times p$ dimensională

$$\frac{\partial g(\theta)}{\partial \theta} = \begin{pmatrix} \dfrac{\partial g_1(\theta)}{\partial \theta_1} & \dfrac{\partial g_1(\theta)}{\partial \theta_2} & \cdots & \dfrac{\partial g_1(\theta)}{\partial \theta_p} \\ \dfrac{\partial g_2(\theta)}{\partial \theta_1} & \dfrac{\partial g_2(\theta)}{\partial \theta_2} & \cdots & \dfrac{\partial g_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_r(\theta)}{\partial \theta_1} & \dfrac{\partial g_r(\theta)}{\partial \theta_2} & \cdots & \dfrac{\partial g_r(\theta)}{\partial \theta_p} \end{pmatrix}.$$

Considerăm acum un curent continuu într-un WGN, ce are parametrii A şi σ^2 necunoscuţi şi ne propunem să estimăm parametrul

$$\alpha := \frac{A^2}{\sigma^2},$$

care poate fi considerat proporţia semnal-zgomot (*SNR, Signal-to-Noise Ratio*). Vectorul parametrilor este

$$\theta = (\mathbf{A}, \sigma^2)^t, \qquad \text{iar} \qquad g(\theta) = \frac{\theta_1^2}{\theta_2} = \frac{A^2}{\sigma^2}.$$

Determinăm, prin calcul imediat:

$$\begin{cases} I(\theta) &= \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \quad \text{iar Jacobianul este} \\ \frac{\partial g(\theta)}{\partial \theta} &= \begin{pmatrix} \frac{\partial g(\theta)}{\partial \theta_1} & \frac{\partial g(\theta)}{\partial \theta_1} \end{pmatrix} = \begin{pmatrix} \frac{\partial g(\theta)}{\partial A} & \frac{\partial g(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{2A}{\sigma^2} & -\frac{A^2}{\sigma^4} \end{pmatrix}, \\ \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} &= \begin{pmatrix} \frac{2A}{\sigma^2} & -\frac{A^2}{\sigma^4} \end{pmatrix} \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} \begin{pmatrix} \frac{2A}{\sigma^2} \\ -\frac{A^2}{\sigma^4} \end{pmatrix} = \frac{4A^2}{n\sigma^2} + \frac{2A^4}{n\sigma^4} = \frac{4\alpha + 2\alpha^2}{n}. \end{cases}$$

Cum α este scalar,

$$D^2(\hat{\alpha}) \geq \frac{4\alpha + 2\alpha^2}{n}.$$

Aşa cum deja am discutat, eficienţa estimatorului se păstrează în urma transformărilor liniare de tipul

$$\alpha = g(\theta) = \mathbf{A}\theta + \mathbf{b}, \quad \mathbf{A} \in \mathcal{M}_{r\times p}(\mathbb{R}), \quad \mathbf{b} \in \mathcal{M}_{r\times 1}(\mathbb{R}).$$

Dacă $\hat{\alpha} = \mathbf{A}\hat{\theta} + \mathbf{b}$, iar $\hat{\theta}$ este un estimator eficient (adică $C_{\hat{\theta}} = I^{-1}(\theta)$), atunci $\mathbb{E}(\hat{\alpha}) = \mathbf{A}\theta + \mathbf{b} = \alpha$ (adică $\hat{\alpha}$ este nedeplasat) şi avem:

$$C_{\hat{\alpha}} = \mathbf{A}C_{\hat{\theta}}\mathbf{A}^t = \mathbf{A}I^{-1}(\theta)\mathbf{A}^t = \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^t}{\partial \theta} = CRLB(\hat{\alpha}).$$

Pentru transformările neliniare, eficienţa este menţinută "aproximativ" doar pentru $n \to +\infty$. Aceasta înseamnă că, pentru un volum de selecţie suficient de mare, densitatea de repartiţie a estimatorului $\hat{\theta}$ este concentrată în jurul valorii reale, teoretice, a parametrului θ, adică estimatorul este consistent.

1.5 CRLB în cazul Gaussian general

În cazul în care observaţiile sunt de tip Gaussian, putem deduce o formulă pentru *CRLB* care să generalizeze (9). Presupunem, pentru aceasta, că

$$V \sim \mathcal{N}(\mu(\theta), C(\theta)), \quad \text{unde} \quad \mu(\theta) \in \mathcal{M}_{n\times 1}(\mathbb{R}) \quad \text{şi} \quad C(\theta) \in \mathcal{M}_{n\times n}(\mathbb{R}).$$

Este evident că media şi dispersia teoretice depind de parametrul θ.

Informaţia Fisher va fi:

$$(I(\theta))_{ij} = \left(\frac{\partial \mu(\theta)}{\partial \theta_i}\right)^t C^{-1}(\theta) \left(\frac{\partial \mu(\theta)}{\partial \theta_j}\right) + \frac{1}{2} \operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_i} C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_j}\right), \quad \text{unde} \tag{25}$$

$$\frac{\partial \mu(\theta)}{\partial \theta_i} = \begin{pmatrix} \frac{\partial (\mu(\theta))_1}{\partial \theta_i} \\ \frac{\partial (\mu(\theta))_2}{\partial \theta_i} \\ \vdots \\ \frac{\partial (\mu(\theta))_n}{\partial \theta_i} \end{pmatrix} \quad \text{şi} \quad \frac{\partial C(\theta)}{\partial \theta_i} = \begin{pmatrix} \frac{\partial (C(\theta))_{11}}{\partial \theta_i} & \frac{\partial (C(\theta))_{12}}{\partial \theta_i} & \cdots & \frac{\partial (C(\theta))_{1n}}{\partial \theta_i} \\ \frac{\partial (C(\theta))_{21}}{\partial \theta_i} & \frac{\partial (C(\theta))_{22}}{\partial \theta_i} & \cdots & \frac{\partial (C(\theta))_{2n}}{\partial \theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial (C(\theta))_{n1}}{\partial \theta_i} & \frac{\partial (C(\theta))_{n2}}{\partial \theta_i} & \cdots & \frac{\partial (C(\theta))_{nn}}{\partial \theta_i} \end{pmatrix}.$$

Pentru situaţia $V \sim \mathcal{N}(\mu(\theta), C(\theta))$, aceasta se reduce la:

$$I(\theta) = \left(\frac{\partial \mu(\theta)}{\partial \theta}\right)^t C^{-1}(\theta) \left(\frac{\partial \mu(\theta)}{\partial \theta}\right) + \frac{1}{2} \operatorname{tr}\left(\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_i}\right)^2\right), \tag{26}$$

după cum vedem în rezultatul tehnic ce urmează.

Densitatea de repartiţie a vectorului aleator de selecţie este:

$$f\left(v,\theta\right)=\frac{1}{\left(2\pi\right)^{n/2}\sqrt{\det\left(C(\theta)\right)}}\exp\left(-\frac{1}{2}\left(v-\mu(\theta)\right)^{t}C^{-1}(\theta)\left(v-\mu(\theta)\right)\right).$$

Vom utiliza următoarele identităţi:

$$\frac{\partial\ln\left(\det\left(C(\theta)\right)\right)}{\partial\theta_k}=\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial\theta_k}\right)\qquad\text{şi}\qquad\frac{\partial C^{-1}(\theta)}{\partial\theta_k}=-C^{-1}(\theta)\frac{\partial C(\theta)}{\partial\theta_k}C^{-1}(\theta),\tag{27}$$

unde $\partial C(\theta)/\partial\theta_k$ este matricea de tipul $n\times n$, al cărei element de pe poziţia (i,j) este $\partial\left(C(\theta)\right)_{ij}/\partial\theta_k$. Pentru a obţine prima formula din (27), observăm, pentru început că

$$\frac{\partial\ln\left(\det\left(C(\theta)\right)\right)}{\partial\theta_k}=\frac{1}{\det\left(C(\theta)\right)}\frac{\partial\det\left(C(\theta)\right)}{\partial\theta_k}.\tag{28}$$

Cum $\det\left(C(\theta)\right)$ depinde de toate elementele matricei $C(\theta)$,

$$\frac{\partial\det\left(C(\theta)\right)}{\partial\theta_k}=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\det\left(C(\theta)\right)}{\partial\left(C(\theta)\right)_{ij}}\frac{\partial\left(C(\theta)\right)_{ij}}{\partial\theta_k}=\operatorname{tr}\left(\frac{\partial\det\left(C(\theta)\right)}{\partial C(\theta)}\frac{\partial C(\theta)^{t}}{\partial\theta_k}\right),\tag{29}$$

unde $\partial\det\left(C(\theta)\right)/\partial C(\theta)$ este o matrice de tip $n\times n$, al cărei element de pe poziţia (i,j) este $\partial\det\left(C(\theta)\right)/\partial\left(C(\theta)\right)_{ij}$. Am ţinut cont şi de identitatea matriceală $\operatorname{tr}\left(AB^{t}\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}B_{ij}$. Din definiţia determinantului,

$$\det\left(C(\theta)\right)=\sum_{i=1}^{n}\left(C(\theta)\right)_{ij}M_{ij},\quad M\in\mathcal{M}_{n\times n}\left(\mathbb{R}\right)\text{ este matricea cofactor şi }j\text{ este arbitrar în }\{1,2,...,n\}.$$

Prin urmare,

$$\frac{\partial\det\left(C(\theta)\right)}{\partial\left(C(\theta)\right)_{ij}}=M_{ij},\quad\text{sau, echivalent,}\quad\frac{\partial\det\left(C(\theta)\right)}{\partial C(\theta)}=M.$$

Pe de altă parte,

$$C^{-1}(\theta)=\frac{M^{t}}{\det\left(C(\theta)\right)}\qquad\text{şi deci}\qquad\frac{\partial\det\left(C(\theta)\right)}{\partial C(\theta)}=C^{-1}(\theta)\det\left(C(\theta)\right)$$

Folosim acum (28) şi (29) şi obţinem rezultatul dorit:

$$\frac{\partial\ln\left(\det\left(C(\theta)\right)\right)}{\partial\theta_k}=\frac{1}{\det\left(C(\theta)\right)}\operatorname{tr}\left(C^{-1}(\theta)\det\left(C(\theta)\right)\frac{\partial C(\theta)}{\partial\theta_k}\right)=\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial\theta_k}\right).$$

Pentru cea de a doua identitate din (27), plecăm de la egalitatea evidentă $C^{-1}(\theta)C(\theta)=I$. Diferenţiind componentele matricelor şi scriind sub formă matriceală, obţinem:

$$C^{-1}(\theta)\frac{\partial C(\theta)}{\partial\theta_k}+\frac{\partial C^{-1}(\theta)}{\partial\theta_k}C(\theta)=\mathbf{0},$$

iar concluzia rezultă imediat.

Ne preocupăm acum de determinarea *CRLB*. Derivatele parţiale în raport cu parametrii conduc la următoarea formă a funcţiei de verosimilitate:

$$\frac{\partial\ln f\left(V,\theta\right)}{\partial\theta_k}=-\frac{1}{2}\frac{\partial\ln\det\left(C(\theta)\right)}{\partial\theta_k}-\frac{1}{2}\frac{\partial}{\partial\theta_k}\left(\left(V-\mu(\theta)\right)^{t}C^{-1}(\theta)\left(V-\mu(\theta)\right)\right).$$

Cum primul termen a fost evaluat în (27), ne vom ocupa de cel de al doilea:

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k}\left(\left(V-\mu(\theta)\right)^{t}C^{-1}(\theta)\left(V-\mu(\theta)\right)\right)=\frac{\partial}{\partial\theta_k}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left(X_i-\left(\mu(\theta)\right)_i\right)\left(C^{-1}(\theta)\right)_{ij}\left(X_j-\left(\mu(\theta)\right)_j\right)\right)\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(X_i-\left(\mu(\theta)\right)_i\right)\left[\left(C^{-1}(\theta)\right)_{ij}\left(-\frac{\partial\left(\mu(\theta)\right)_j}{\partial\theta_k}\right)+\frac{\partial\left(C^{-1}(\theta)\right)_{ij}}{\partial\theta_k}\left(X_j-\left(\mu(\theta)\right)_j\right)\right]\right.\\
&\qquad\left.+\left(-\frac{\partial\left(\mu(\theta)\right)_i}{\partial\theta_k}\right)\left(C^{-1}(\theta)\right)_{ij}\left(X_j-\left(\mu(\theta)\right)_j\right)\right\}\\
&=-\left(V-\mu(\theta)\right)^{t}C^{-1}(\theta)\frac{\partial\mu(\theta)}{\partial\theta_k}+\left(V-\mu(\theta)\right)^{t}\frac{\partial C^{-1}(\theta)}{\partial\theta_k}\left(V-\mu(\theta)\right)-\frac{\partial\mu(\theta)^{t}}{\partial\theta_k}C^{-1}(\theta)\left(V-\mu(\theta)\right)\\
&=-2\frac{\partial\mu(\theta)^{t}}{\partial\theta_k}C^{-1}(\theta)\left(V-\mu(\theta)\right)+\left(V-\mu(\theta)\right)^{t}\frac{\partial C^{-1}(\theta)}{\partial\theta_k}\left(V-\mu(\theta)\right).
\end{aligned}$$

Folosind (27) împreună cu precedentul rezultat obţinem:

$$\frac{\partial \ln f(V,\theta)}{\partial \theta_k} = -\frac{1}{2}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right) + \frac{\partial \mu(\theta)^t}{\partial \theta_k} C^{-1}(\theta)\left(V-\mu(\theta)\right) - \frac{1}{2}\left(V-\mu(\theta)\right)^t \frac{\partial C^{-1}(\theta)}{\partial \theta_k}\left(V-\mu(\theta)\right). \quad (30)$$

Notăm acum $Y := V - \mu(\theta)$. Pentru a determina *CRLB* trebuie să evaluăm elementele

$$(I(\theta))_{kl} = \mathbb{E}\left(\frac{\partial \ln f(V,\theta)}{\partial \theta_k}\frac{\partial \ln f(V,\theta)}{\partial \theta_l}\right), \quad k,l \in \{1,2,...,p\}.$$

Avem, astfel,

$$\begin{aligned}(I(\theta))_{kl} &= \frac{1}{4}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right)\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_l}\right) + \frac{1}{2}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right)\mathbb{E}\left(Y^t\frac{\partial C^{-1}(\theta)}{\partial \theta_l}Y\right)\\ &+\frac{\partial \mu(\theta)^t}{\partial \theta_k}C^{-1}(\theta)\mathbb{E}\left(YY^t\right)C^{-1}(\theta)\frac{\partial \mu(\theta)}{\partial \theta_l} + \frac{1}{4}\mathbb{E}\left(Y^t\frac{\partial C^{-1}(\theta)}{\partial \theta_k}YY^t\frac{\partial C^{-1}(\theta)}{\partial \theta_l}Y\right),\end{aligned}$$

toate momentele de ordin impar fiind zero. Continuând evaluarea,

$$\begin{aligned}(I(\theta))_{kl} &= \frac{1}{4}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right)\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_l}\right) - \frac{1}{2}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right)\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_l}\right)\\ &+\frac{\partial \mu(\theta)^t}{\partial \theta_k}C^{-1}(\theta)\frac{\partial \mu(\theta)}{\partial \theta_l} + \frac{1}{4}\mathbb{E}\left(Y^t\frac{\partial C^{-1}(\theta)}{\partial \theta_k}YY^t\frac{\partial C^{-1}(\theta)}{\partial \theta_l}Y\right).\end{aligned} \quad (31)$$

Am utilizat (27) şi faptul că $\mathbb{E}\left(Y^tZ\right) = \operatorname{tr}\left(\mathbb{E}\left(ZY^t\right)\right)$, pentru $Y, Z \in \mathcal{M}_{n\times 1}(\mathbb{R})$. Pentru estimarea ultimului termen, folosim un rezultat furnizat de Porat, Friedlander, 1986:

$$\mathbb{E}\left(Y^tAYY^tBY\right) = \operatorname{tr}(AC)\operatorname{tr}(BC) + 2\operatorname{tr}(ACBC), \quad \text{unde } C := \mathbb{E}\left(YY^t\right),$$

iar matricele A şi B sunt simetrice. În virtutea relaţiei anterioare şi, folosind din nou relaţia (27), ultimul termen din reprezentarea (31) devine:

$$\begin{aligned}&\frac{1}{4}\operatorname{tr}\left(\frac{\partial C^{-1}(\theta)}{\partial \theta_k}C(\theta)\right)\operatorname{tr}\left(\frac{\partial C^{-1}(\theta)}{\partial \theta_l}C(\theta)\right) + \frac{1}{2}\operatorname{tr}\left(\frac{\partial C^{-1}(\theta)}{\partial \theta_k}C(\theta)\frac{\partial C^{-1}(\theta)}{\partial \theta_l}C(\theta)\right)\\ &= \frac{1}{4}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}\right)\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_l}\right) + \frac{1}{2}\operatorname{tr}\left(C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_k}C^{-1}(\theta)\frac{\partial C(\theta)}{\partial \theta_l}\right).\end{aligned} \quad (32)$$

Inserăm (32) în (31) şi obţinem rezultatul dorit.

Exemplul 1.7 ***Parametrii unui semnal în WGN.***

I. *Presupunem că dorim să estimăm un parametru scalar σ^2 pentru setul de date*

$$X_i = s_i(\theta) + W_i, \quad \textit{unde } W_i \textit{ este WGN}, \quad i \in \{0,1,...,n-1\}.$$

Matricea de covarianţă este $C = -I$ şi este independentă de parametrul θ, ceea ce face ca al doilea termen din (26) să fie zero. Obţinem astfel:

$$I(\theta) = \frac{1}{\sigma^2}\left(\left(\frac{\partial \mu(\theta)}{\partial \theta}\right)^t\left(\frac{\partial \mu(\theta)}{\partial \theta}\right)\right) = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}\left(\frac{\partial\left(\mu(\theta)\right)_i}{\partial \theta}\right)^2 = \frac{1}{\sigma^2}\sum_{i=0}^{n-1}\left(\frac{\partial s_i(\theta)}{\partial \theta}\right)^2,$$

ceea ce este în concordanţă cu (9). Dacă dorim să generalizăm la cazul unui parametru vectorial al semnalului, din (25) obţinem:

$$(I(\theta))_{ij} = \frac{1}{\sigma^2}\sum_{k=0}^{n-1}\frac{\partial s_k(\theta)}{\partial \theta_i}\frac{\partial s_k(\theta)}{\partial \theta_j}.$$

II. *Dacă observăm $X_i = W_i$, unde W_i este WGN cu dispersia $\theta = \sigma^2$, $i \in \{0,1,...,n-1\}$, atunci, în conformitate cu (26), deoarece $C(\sigma^2) = \sigma^2 I$, avem:*

$$I(\theta) = \frac{1}{2}\operatorname{tr}\left(\left(C^{-1}(\sigma^2)\frac{\partial C(\sigma^2)}{\partial \sigma^2}\right)^2\right) = \frac{1}{2}\operatorname{tr}\left(\left(\frac{1}{\sigma^2}I\right)^2\right) = \frac{1}{2\sigma^4}\operatorname{tr}(I) = \frac{n}{2\sigma^4},$$

ceea ce este în concordanţă cu Exemplul 1.5.

***III**. Presupunem datele:*

$$X_i = A + W_i, \quad \textit{unde } W_i \textit{ este WGN}, \quad i \in \{0, 1, ..., n-1\},$$

iar A, mărimea (amplitudinea) curentului continuu, este o variabilă aleatoare Gaussiană, de medie 0 şi dispersie σ_A^2 ($A \sim \mathcal{N}\left(0, \sigma_A^2\right)$), independentă de $i \in \{0, 1, ..., n-1\}$. Puterea semnalului (dispersia σ_A^2) se presupune a fi parametrul necunoscut. Vectorul de selecţie $V = (X_0, X_1, ..., X_{n-1})^t$ va fi, prin urmare, un vector repartizat tot normal, de medie $\mathbf{0}$ şi având matricea de covarianţă de tipul $n \times n$, ale cărei elemente sunt:

$$\left(C(\sigma_A^2)\right)_{ij} = \mathbb{E}\left(X_i X_j\right) = \mathbb{E}\left(\left(A + W_i\right)\left(A + W_j\right)\right) = \sigma_A^2 + \sigma^2 \delta_{ij}.$$

Aceasta înseamnă, sub formă vectorială,

$$C(\sigma_A^2) = \sigma_A^2 \mathbf{1} \cdot \mathbf{1}^t + \sigma^2 I, \qquad \textit{unde} \qquad \mathbf{1} = (1, 1, ..., 1)^t \in \mathcal{M}_{n\times 1}\left(\mathbb{R}\right).$$

Aplicăm acum identitatea Woodbury privind inversa unei matrice având o anumită structură particulară, dată (vezi Kay [2, Annex A1.1.3, Matrix Manipulation and Formulas]): dacă $A \in \mathcal{M}_{n\times n}\left(\mathbb{R}\right)$ şi $u \in \mathcal{M}_{n\times 1}\left(\mathbb{R}\right)$, atunci are loc relaţia:

$$\left(A + uu^t\right)^{-1} = A^{-1} - \frac{A^{-1} u u^t A^{-1}}{1 + u^t A^{-1} u}.$$

Obţinem astfel:

$$C^{-1}(\sigma_A^2) = \frac{1}{\sigma^2}\left(I - \frac{\sigma_A^2}{\sigma^2 + n\sigma_A^2} \mathbf{1} \cdot \mathbf{1}^t\right).$$

Deoarece

$$\frac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = \mathbf{1} \cdot \mathbf{1}^t, \quad \textit{deducem} \quad C^{-1}(\sigma_A^2) \frac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = \frac{1}{\sigma^2 + n\sigma_A^2} \mathbf{1} \cdot \mathbf{1}^t.$$

Înlocuim în (26) şi avem:

$$I(\sigma_A^2) = \frac{1}{2} \operatorname{tr}\left(\left(\frac{1}{\sigma^2 + n\sigma_A^2}\right)^2 \mathbf{1} \cdot \mathbf{1}^t \cdot \mathbf{1} \cdot \mathbf{1}^t\right) = \frac{n}{2}\left(\frac{1}{\sigma^2 + n\sigma_A^2}\right)^2 \operatorname{tr}\left(\mathbf{1} \cdot \mathbf{1}^t\right) = \frac{1}{2}\left(\frac{n}{\sigma^2 + n\sigma_A^2}\right)^2,$$

adică CRLB este

$$D^2\left(\widehat{\sigma_A^2}\right) \geq 2\left(\sigma_A^2 + \frac{\sigma^2}{n}\right)^2 = CRLB\left(\widehat{\sigma_A^2}\right).$$

Observăm că, chiar şi pentru $n \to +\infty$, $CRLB\left(\widehat{\sigma_A^2}\right)$ obţinut nu poate descreşte sub valoarea $2\sigma_A^4$. Aceasta se întâmplă datorită faptului că fiecare dată suplimentară în eşantion vine cu aceeaşi valoare pentru A.

1.6 CRLB asimptotică pentru procese aleatoare Gaussiene staţionare în sens larg

Uneori, în practică, pote fi mai dificil de determinat *CRLB* folosind (25) datorită necesităţii determinării inversei matricei de covarianţă. Deşi aceasta se poate determina numeric, folosind un software matematic, există şi o abordare alternativă, dar care poate fi aplicată doar proceselor Gaussiene staţionare în sens larg (*WSS, Wide Sense Stationary*). Pentru aceasta, volumul de selecţie $n \to +\infty$. Pentru procese cu densitatea spectrală de putere (*PSD, Power Spectral Density*) largă, aproximarea va fi bună pentru înregistrări de date de lungime moderată, în timp ce procese cu bandă îngustă trebuiesc înregistrări de date de lungime sporită.

Elementele matricei informaţiei Fisher sunt, asimptotic, pentru $n \to +\infty$,

$$\left(I(\theta)\right)_{ij} = \frac{n}{2} \int_{-1/2}^{1/2} \frac{\partial \ln P_{xx}\left(f, \theta\right)}{\partial \theta_i} \frac{\partial \ln P_{xx}\left(f, \theta\right)}{\partial \theta_j} df, \tag{33}$$

unde $P_{xx}\left(f, \theta\right)$ este indicatorul *PSD* al procesului, cu dependenţa explicită de parametrul vectorial θ (vezi Kay [2, *Appendix 3D*]). Presupunem că $\mathbb{E}\left(X_i\right) = 0$, pentru fiecare i.

Exemplul 1.8 ***Frecvenţa centrală a procesului.*** *O problemă uzuală constă în estimarea frecvenţei centrale f_c a unei PSD, care, de altfel, este cunoscută. Presupunem că este dată:*

$$P_{xx}(f, f_c) = Q\left(f - f_c\right) + Q\left(-f - f_c\right) + \sigma^2$$

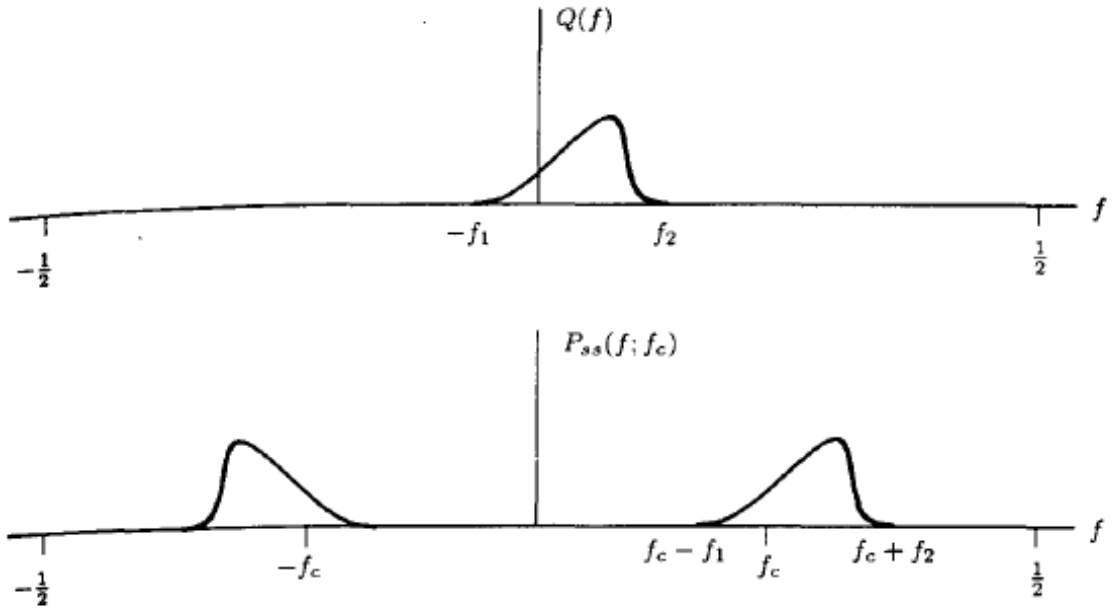


Figura 2. $f_{c_{\min}} = f_1$ $f_{c_{\max}} = \frac{1}{2} - f_2$

şi dorim să determinăm CRLB pentru f_c, *în ipoteza că* $Q(f)$ *şi* σ^2 *sunt cunoscute. Putem percepe procesul ca fiind un semnal aleator aflat într-un WGN. Frecvenţele centrale posibile sunt constrânse a fi situate în intervalul* $[f_1, 1/2 - f_2]$. *Pentru aceste frecvenţe centrale, PSD-ul semnalului, pentru* $f \geq 0$, *se va situa în intervalul* $[0, 1/2]$. *Atunci, cum* $\theta = f_c$ *este un scalar, din (33), obţinem:*

$$D^2(\hat{f}_c) \geq \frac{1}{\frac{n}{2}\int_{-1/2}^{1/2}\left(\frac{\partial \ln P_{xx}(f,f_c)}{\partial f_c}\right)^2 df}.$$

Dar,

$$\frac{\partial \ln P_{xx}(f,f_c)}{\partial f_c} = \frac{\partial \ln\left(Q(f-f_c)+Q(-f-f_c)+\sigma^2\right)}{\partial f_c} = \frac{\frac{\partial Q(f-f_c)}{\partial f_c}+\frac{\partial Q(-f-f_c)}{\partial f_c}}{Q(f-f_c)+Q(-f-f_c)+\sigma^2},$$

care, fiind o funcţie impară în raport cu f, *conduce la:*

$$\int_{-1/2}^{1/2}\left(\frac{\partial \ln P_{xx}(f,f_c)}{\partial f_c}\right)^2 df = 2\int_0^{1/2}\left(\frac{\partial \ln P_{xx}(f,f_c)}{\partial f_c}\right)^2 df.$$

De asemenea, pentru $f \geq 0$, $Q(-f-f_c) = 0$, *iar derivata sa va fi egală cu zero în virtutea celor discutate privind restricţiile. Rezultă că:*

$$\begin{aligned} D^2(\hat{f}_c) &\geq \frac{1}{n\int_0^{1/2}\left(\frac{\partial Q(f-f_c)/\partial f_c}{Q(f-f_c)+\sigma^2}\right)^2 df} = \frac{1}{n\int_0^{1/2}\left(\frac{-\partial Q(f-f_c)/\partial (f-f_c)}{Q(f-f_c)+\sigma^2}\right)^2 df} \\ &= \frac{1}{n\int_{-f_c}^{1/2-f_c}\left(\frac{-\partial Q(f')/\partial f'}{Q(f')+\sigma^2}\right)^2 df'}, \quad \text{unde} \quad f' = f - f_c. \end{aligned}$$

Dar $1/2 - f_c \geq 1/2 - f_{c_{\max}} = f_2$ *şi* $-f_c \leq -f_{c_{\min}} = -f_1$, *putând astfel schimba capetele de integrare în intervalul* $[-1/2, 1/2]$. *Obţinem:*

$$D^2(\hat{f}_c) \geq \frac{1}{n\int_{-1/2}^{1/2}\left(\frac{\partial Q(f)/\partial f}{Q(f)+\sigma^2}\right)^2 df} = \frac{1}{n\int_{-1/2}^{1/2}\left(\frac{\partial \ln\left(Q(f)+\sigma^2\right)}{\partial f}\right)^2 df}.$$

Dacă considerăm, ca un caz particular,

$$Q(f) := \exp\left(-\frac{1}{2}\left(\frac{f}{\sigma_f}\right)^2\right), \quad \textit{unde } \sigma_f \ll 1/2,$$

atunci $Q(f) \gg \sigma^2$ şi avem, asimptotic,

$$D^2(\hat{f}_c) \geq \frac{1}{n\int_{-1/2}^{1/2} \frac{f^2}{\sigma_f^4} df} = \frac{12\sigma_f^4}{n}.$$

Lăţimi de bandă mai mici decât σ_f^4 vor conduce la limite inferioare de mărginire mai joase pentru frecvenţa centrală deoarece PSD-ul se modifică mai rapid odată cu schimbările lui f_c.

1.6.1 Exemple privind procesarea semnalelor

Aplicăm teoria deducerii *CRLB* pentru câteva probleme de interes privind procesarea semnalelor: *estimarea ariei de acoperire, estimarea poziţiei ţintei* (sonar, radar, robotică), *estimarea frecvenţei* (sonar, radar, econometrie, spectrometrie), *estimarea parametrilor autoregresivi* (teoria limbajului, econometrie).

Exemplul 1.9 ***Estimarea range-ului (raza de acţiune).*** *Un radar emite un semnal de tip puls. Timpul total al traseului semnalului de la emiţător la ţintă şi înapoi va fi notat cu τ_0 şi este dependent de range (R) prin formula $\tau_0 = 2R/c$, unde c este viteza sa de propagare. Estimarea range-ului este deci echivalentă cu estimarea timpului de întoarcere a semnalului la sursa emitentă. Dacă $s(t)$ este semnalul transmis, un model simplu pentru unda continuă recepţionată este:*

$$x(t) = s(t - \tau_0) + W(t), \quad t \in [0, T].$$

Presupunem că pulsul transmis este diferit de zero pe $[0, T_s]$ şi este limitat la o lăţime de bandă de B Hz. Dacă întârzierea de timp până la recepţionarea semnalului revenit este $\tau_{0_{\max}}$, atunci intervalul de observare trebuie ales astfel încât să includă întregul semnal, adică $T = T_s + \tau_{0_{\max}}$. Zgomotul pe sistem este de tip Gaussian, cu PSD-ul şi autocorelaţia (ACF-ul - corelaţia unui semnal cu o copie întârziată a sa, ca funcţie de întârziere. Informal, este similaritatea dintre observaţii, în funcţie de intervalul de timp dintre ele) ca în Figura 3.

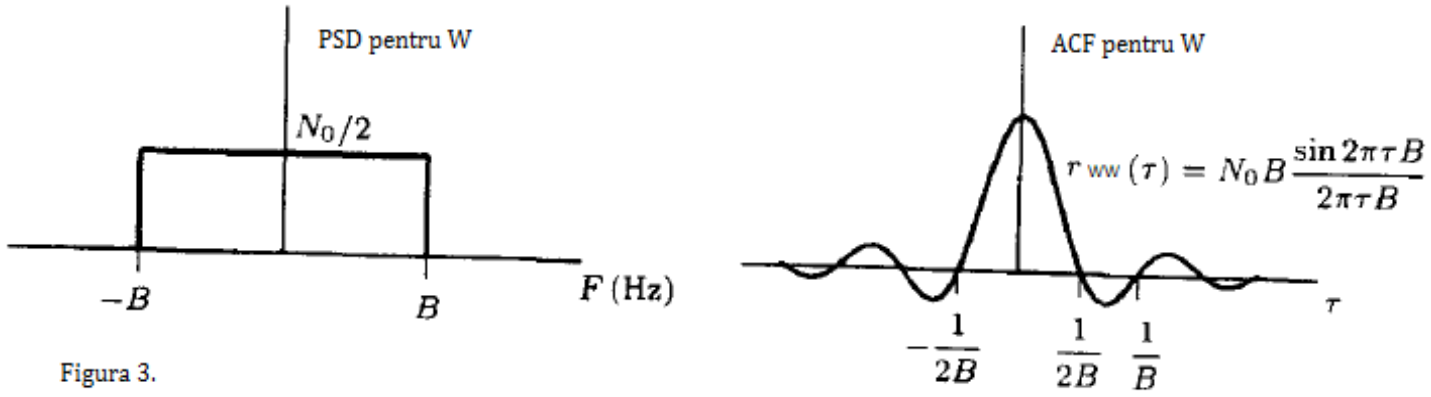


Figura 3.

Limitarea de bandă a zgomotului rezultă prin filtrarea undei la lăţimea de bandă a semnalului de B Hz. Pentru eşantioanare, considerăm $\Delta := 1/(2B)$, iar datele observate vor fi:

$$X_{i\Delta} = s(i\Delta - \tau_0) + W(i\Delta), \quad i \in \{0, 1, ..., n-1\}.$$

Notăm cu X_i şi W_i elementele selecţiei şi construim şirul de date discret:

$$X_i = s(i\Delta - \tau_0) + W_i. \tag{34}$$

Menţionăm că, pentru fiecare i, W_i sunt WGN deoarece observaţiile sunt separate de $k\Delta = k/(2B)$, care corespunde zerourilor funcţiei ACF a lui $W(t)$, după cum se observă în Figura 3. De asemenea,

$$D^2(W_i) = \sigma^2 (= r_{WW}(0)) = N_0 B.$$

Deoarece semnalul este nenul doar pe intervalul $\tau_0 \leq t \leq \tau_0 + T_s$, modelul (34) se reduce la:

$$X_i = \begin{cases} W_i, & 1 \leq i \leq i_0 - 1, \\ s(i\Delta - \tau_0) + W_i, & i_0 \leq i \leq i_0 + M - 1, \\ W_i, & i_0 + M \leq i \leq n, \end{cases} \tag{35}$$

unde M este lungimea semnalului eşantionat, iar $i_0 = \tau_0/\Delta$ reprezintă întârzierea, în eşantioane. Pentru simplitate, se presupune Δ mic astfel încât τ_0/Δ să poată fi aproximat cu un întreg. Cu această formulare, putem aplica (9) pentru evaluarea CRLB:

$$D^2(\hat{\tau}_0) \geq \frac{\sigma^2}{\sum_{i=0}^{n-1}\left(\frac{\partial s(i,\tau_0)}{\partial \tau_0}\right)^2} = \frac{\sigma^2}{\sum_{i=i_0}^{i_0+M-1}\left(\frac{\partial s(i\Delta-\tau_0)}{\partial \tau_0}\right)^2} = \frac{\sigma^2}{\sum_{i=i_0}^{i_0+M-1}\left(\frac{ds(t)}{dt}|_{t=i\Delta-\tau_0}\right)^2} = \frac{\sigma^2}{\sum_{i=0}^{M-1}\left(\frac{ds(t)}{dt}|_{t=i\Delta}\right)^2},$$

deoarece $\tau_0 = i_0\Delta$. Cum Δ este suficient de mic, aproximăm suma cu o integrală şi obţinem:

$$D^2(\hat{\tau}_0) \geq \frac{\sigma^2}{\frac{1}{\Delta}\int_0^{T_s}\left(\frac{ds(t)}{dt}\right)^2 dt} = \frac{N_0/2}{\int_0^{T_s}(s'(t))^2\,dt}, \quad \textit{deoarece} \quad \Delta = 1/(2B) \ \textit{şi} \ \sigma^2 = N_0 B. \tag{36}$$

Cum energia poate fi reprezentată integral, $\mathcal{E} = \int_0^{T_s} s^2(t)\,dt$, inegalitatea (36) se poate scrie, sub formă echivalentă,

$$D^2(\hat{\tau}_0) \geq \frac{1}{\frac{\mathcal{E}}{N_0/2}\overline{F^2}}, \quad \textit{unde} \quad \overline{F^2} = \frac{\int_0^{T_s}(s'(t))^2\,dt}{\int_0^{T_s} s^2(t)\,dt}.$$

$\overline{F^2}$ este o măsură a lăţimii de bandă a semnalului. Cu cât este mai mare această cantitate, cu atât mai mică va fi CRLB.

Exemplul 1.10 ***Estimarea parametrului vectorial al semnalului sinusoidal.*** *În multe aplicaţii se pune problema estimării parametrilor unui semnal de tip sinusoidal. Sonarele şi mecanismele de tip radar pot cataloga semnalul observat ca fiind unul de tip sinusoidal. Ne punem problema determinării CRLB pentru amplitudinea A, frecvenţa f_0 şi faza ϕ a unui curent sinusoidal într-un WGN. Datele sunt de tipul:*

$$X_i = A\cos(2\pi f_0 i + \phi) + W_i, \quad i = 0, 1, ..., n-1,$$

unde $A > 0$ şi $0 < f_0 < 1/2$. Deoarece avem mai mulţi parametri necunoscuţi, folosim formula:

$$(I(\theta))_{ij} = \frac{1}{\sigma^2}\sum_{k=0}^{n-1}\frac{\partial s_k(\theta)}{\partial \theta_i}\frac{\partial s_k(\theta)}{\partial \theta_j}, \quad \textit{pentru} \quad \theta = (A, f_0, \phi)^t.$$

Pentru evaluarea CRLB, presupunem că f_0 nu este în vecinătatea lui 0 sau $1/2$, fapt ce ne permite să utilizăm aproximările următoare (vezi Stoica, P.; R.L. Moses; B. Friedlander; T. Soderstrom, Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids from Noisy Measurements, IEEE Trans. Acoust., Speech, Signal Process., Vol. 37, pp. 378-392, March 1989):

$$\frac{1}{n^{i+1}}\sum_{k=0}^{n-1} k^i \sin(4\pi f_0 k + 2\phi) \approx 0 \qquad \textit{şi} \qquad \frac{1}{n^{i+1}}\sum_{k=0}^{n-1} k^i \cos(4\pi f_0 k + 2\phi) \approx 0, \quad i \in \{0, 1, 2\}.$$

Folosind aceste aproximări şi notând $\alpha := 2\pi f_0 i + \phi$, *avem:*

$$\begin{cases} (I(\theta))_{11} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}\cos^2\alpha = \dfrac{1}{\sigma^2}\sum_{k=0}^{n-1}\left(\dfrac{1}{2}+\dfrac{1}{2}\cos 2\alpha\right) \approx \dfrac{n}{2\sigma^2}, \\ (I(\theta))_{12} &= -\dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}2A\pi k\cos\alpha\sin\alpha = -\dfrac{\pi A}{\sigma^2}k\sin 2\alpha \approx 0, \\ (I(\theta))_{13} &= -\dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}A\cos\alpha\sin\alpha = -\dfrac{A}{2\sigma^2}\sum_{k=0}^{n-1}\sin 2\alpha \approx 0, \\ (I(\theta))_{22} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}A^2(2\pi k)^2\sin^2\alpha = \dfrac{(2\pi A)^2}{\sigma^2}\sum_{k=0}^{n-1}k^2\left(\dfrac{1}{2}-\dfrac{1}{2}\cos 2\alpha\right) \approx \dfrac{(2\pi A)^2}{2\sigma^2}\sum_{k=0}^{n-1}k^2, \\ (I(\theta))_{23} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}A^2 2\pi k\sin^2\alpha \approx \dfrac{\pi A^2}{\sigma^2}\sum_{k=0}^{n-1}k, \\ (I(\theta))_{33} &= \dfrac{1}{\sigma^2}\displaystyle\sum_{k=0}^{n-1}A^2\sin^2\alpha \approx \dfrac{nA^2}{2\sigma^2}. \end{cases}$$

Matricea informaţiei Fisher este:

$$I(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \dfrac{n}{2} & 0 & 0 \\ 0 & 2\pi^2A^2\displaystyle\sum_{k=0}^{n-1}k^2 & \pi A^2\displaystyle\sum_{k=0}^{n-1}k \\ 0 & \pi A^2\displaystyle\sum_{k=0}^{n-1}k & \dfrac{nA^2}{2} \end{pmatrix}.$$

Obţinem astfel marginile inferioare:

$$D^2(\hat{A}) \geq \frac{2\sigma^2}{n}, \qquad D^2(\hat{f}_0) \geq \frac{12}{(2\pi)^2\eta n\left(n^2-1\right)} \qquad şi \qquad D^2(\hat{\phi}) \geq \frac{2\left(2n-1\right)}{\eta n\left(n+1\right)}, \tag{37}$$

unde $\eta := A^2/\left(2\sigma^2\right)$ *este coeficientul SNR (signal-to-noise ratio) al semnalului.*

Exemplul 1.11 ***Estimarea menţinerii ţintei semnalului.*** *În problemele de localizare, este importantă estimarea modului în care este identificată poziţia unei ţinte. Presupunem câmpul acustic de presiune fiind format dintr-un şir de senzori plasaţi echidistant şi că ţinta radiază un semnal de tip sinusoidal de forma* $A\cos\left(2\pi F_0 t+\phi\right)$. *Semnalul recepţionat de senzorul cu indicele* i *este* $A\cos\left(2\pi F_0\left(t-t_i\right)+\phi\right)$, *unde* t_i *este timpul de propagare până la senzorul indicele* i. *Aşa cum se observă în Figura 4, distanţa temporală a frontul de undă la senzorul marcat cu* $i-1$ *diferă faţă de cea de la senzorul marcat cu* i *prin* $d\cos\beta/c$, *datorită distanţei suplimentare de propagare.*

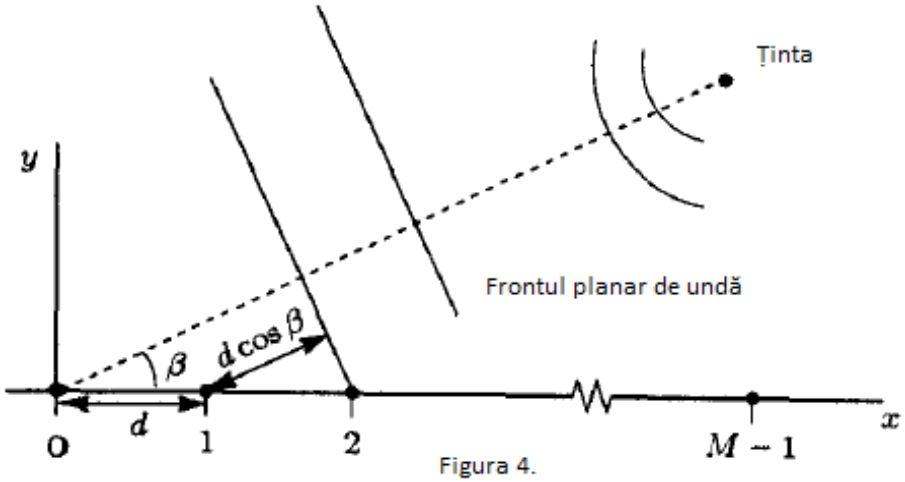


Figura 4.

Prin urmare, timpul de propagare până la senzorul i este:

$$t_i = t_0 - i\frac{d}{c}\cos\beta, \quad i \in \{0, 1, ..., M-1\},$$

unde t_0 este timpul de propagare până la senzorul cu indicele 0. Semnalul observat la senzorul cu indicele i este

$$s_i(t) = A\cos\left(2\pi F_0\left(t - t_0 + i\frac{d}{c}\cos\beta\right) + \phi\right).$$

Dacă se realizează o singură "fotografie" a datelor, adică ne interesează doar informaţia de la un moment fixat t_s,

$$s_i(t_s) = A\cos\left(2\pi\left(F_0\frac{d}{c}\cos\beta\right)i + \phi'\right), \quad \textit{unde} \quad \phi' = \phi + 2\pi F_0(t_s - t_0). \tag{38}$$

Sub această formă, este clar că obsevaţiile spaţiale sunt de tip sinusoidal, cu frecvenţa $f_s = F_0(d/c)\cos\beta$. Presupunem, de asemenea, că datele furnizate de senzor sunt perturbate de câte un zgomot Gaussian de medie 0 şi dispersie σ^2, zgomote ce sunt independente stochastic de la un senzor la altul. Datele eşantionului vor fi modelate prin variabilele de selecţie:

$$X_i = s_i(t_s) + W_i, \quad Wi \textit{ este WGN}, \quad \textit{pentru } i \in \{0, 1, 2, ..., M-1\}.$$

Deoarece, de regulă, A, ϕ, β sunt parametri necunoscuţi, se pune problema estimării parametrilor $\{A, f_s, \phi'\}$ din (38), ca în Exemplul 1.10. Odată ce am determinat CRLB pentru aceştia, putem folosi formulele pentru transformarea parametrilor. Transformarea este $\theta = (A, f_s, \phi')^t$:

$$\alpha = g(\theta) = \begin{pmatrix} A \\ \beta \\ \phi' \end{pmatrix} = \begin{pmatrix} A \\ \arccos\left(\frac{cf_s}{F_0 d}\right) \\ \phi' \end{pmatrix}, \quad \textit{cu Jacobianul} \quad \frac{\partial g(\theta)}{\partial\theta} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{c}{F_0 d\sin\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

În conformitate cu (24), avem:

$$\left(C_{\hat{\alpha}} - \frac{\partial g(\theta)}{\partial\theta} I^{-1}(\theta) \frac{\partial g(\theta)}{\partial\theta}^t\right)_{22} \geq 0.$$

Forma diagonală a Jacobianului implică:

$$D^2(\hat{\beta}) \geq \left(\frac{\partial g(\theta)}{\partial\theta}\right)_{22}^2 \left(I^{-1}(\theta)\right)_{22}.$$

Fromula (37) afirmă că $\left(I^{-1}(\theta)\right)_{22} = 12/\left[(2\pi)^2 \eta M\left(M^2-1\right)\right]$ şi, prin urmare,

$$D^2(\hat{\beta}) \geq \frac{12}{(2\pi)^2 \eta M(M^2-1)} \frac{c^2}{F_0^2 d^2 \sin^2\beta} = \frac{12}{(2\pi)^2 \eta M \frac{M+1}{M-1}\left(\frac{L}{\lambda}\right)^2 \sin^2\beta}, \tag{39}$$

unde $\lambda = c/F_0$ este lungimea de undă a valului propagant planar şi $L = (M-1)d$ este lungimea şirului de senzori. Estimarea este cel mai uşor de realizat dacă $\beta = \pi/2$ şi imposibil de realizat dacă $\beta = 0$.

2 Estimatori nedeplasaţi de varianţă minimă (estimatori MVU)

2.1 Introducere

Am observat că, în evaluarea *CRLB*, obţinem uneori un estimator eficient şi, în consecinţă, acesta este nedeplasat şi de dispersie (varianţă) minimă. Dacă, în schimb, nu există un estimator eficient, se poate formula totuşi problema existenţei şi identificării unui estimator *MVU*. Pentru a realiza acest obiectiv, introducem conceptul de *statistici suficiente* şi, ca instrument de lucru, *Teorema Rao-Blackwell-Lehmann-Scheffe*. Având la dispoziţie această teorie, este posibil, în multe situaţii, determinarea unui estimator *MVU* prin simpla studiere a densităţii de repartiţie a caracteristicii studiate. Vom defini noţiunea de statistică suficientă ca fiind o statistică care sumarizează datele, în sensul că densitatea de repartiţie a datelor nu mai depinde de parametrul necunoscut. Teorema de factorizare Neyman-Fisher ne va permite să găsim statistici suficiente prin examinarea densităţii de repartiţie. Estimatorul *MVU* se poate determina prin găsirea unei funcţii având ca argument statistica suficientă, funcţie ce este un estimator nedeplasat. Pentru utilizarea apoi a Teoremei Rao-Blackwell-Lehmann-Scheffe trebuie să asigurăm faptul că statistica suficientă este completă, în sensul că există o unică funcţie avand-o ca argument ce este estimator nedeplasat.

2.2 Statistici suficiente

Problema estimării nivelului A a unui curent continuu situat într-un WGN revine la a arăta că media de selecţie

$$\hat{A} = \bar{X} = \frac{1}{n}\sum_{i=0}^{n-1} X_i$$

este un estimator *MVU*, de dispersie σ^2/n. Dacă, pe de altă parte, am fi ales $\hat{A} = X_0$ drept estimatorul căutat, este evident faptul că, deşi $\hat{A}$ este nedeplasat, dispersia sa este mult mai mare (este σ^2) decât minimul precedent. Intuitiv, performanţa sa slabă este un reultat al renunţării la variabilele de selecţie $X_1, X_2, ..., X_{n-1}$, ale căror realizări înmagazinează informaţii privind parametrul A. Se poate formula astfel întrebarea "care eşantioane sunt relevante pentru problema estimării". Următoarele mulţimi de date pot fi considerate a fi *suficiente* pentru determinarea estimatorului $\hat{A}$:

$$\begin{cases} S_1 &= \{x_0, x_1, ...x_{n-1}\} \\ S_2 &= \{x_0 + x_1, x_2, x_3, ..., x_{n-1}\} \\ S_3 &= \left\{\sum_{i=0}^{n-1} x_i\right\}, \end{cases}$$

unde, pentru fiecare i, x_i sunt datele observate, corespunzătoare variabilei de selecţie X_i. Mulţimea S_1 reprezintă mulţimea originală de date, care, aşa cum este de aşteptat, este suficientă pentru problema estimării. Cum şi S_2 şi S_3 sunt şi ele suficiente, arată că există mai multe mulţimi suficiente de date. Acea mulţime care conţine numărul minim de elemente se numeşte *mulţimea minimală*. Dacă privim acum elementele acestor mulţimi drept statistici, spunem că cele n statistici ale mulţimii S_1 sunt suficiente, la fel şi cele $n-1$ statistici ale mulţimii S_2, la fel şi unica statistică a mulţimii S_3. Această unică statistică, $\sum_{i=0}^{n-1} X_i$, pe lângă faptul că este statistică suficientă, este chiar *statistica suficientă minimală*. Pentru estimarea parametrului A, odată ce ştim $\sum_{i=0}^{n-1} X_i$, nu mai avem nevoie de datele (valorile) individuale ale fiecărei variabile de selecţie, deoarece toate informaţiile au fost încapsulate în statistica suficientă. Pentru a înţelege ce am spus prin afirmaţia anterioară, considerăm densitatea vectorului aleator de selecţie (adică funcţia de verosimilitate):

$$\mathcal{L}(v, A) = f(v, A) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(x_i - A)^2\right) \tag{40}$$

şi presupunem că au fost observate datele $T(v) = \sum_{i=0}^{n-1} x_i = T_0$. Cunoaşterea valorii statisticii $\sum_{i=0}^{n-1} X_i$ va schimba densitatea de repartiţie în cea condiţionată:

$$f\left(v | \sum_{i=0}^{n-1} x_i = T_0, A\right),$$

care acum furnizează densitatea de repartiţie a observaţiilor după ce statistica suficientă a fost observată. Deoarece statistica este suficientă pentru estimarea parametrului A, această densitate condiţionată ar trebui să nu depindă de parametrul estimat, A.Dacă ar depinde, ar trebui informaţii suplimentare asupra lui A.

Exemplul 2.1 ***Verificarea unei statistici suficiente.*** *Considerăm densitatea vectorială de repartiţie dată de (40). Pentru a arăta că $\sum_{i=0}^{n-1} X_i$ este o statistică suficientă, trebuie să determinăm $f(v|T(v) = T_0, A)$, unde $T(v) = \sum_{i=0}^{n-1} x_i$. Din definiţia densităţii condiţionate, avem:*

$$f(v|T(v) = T_0, A) = \frac{f(v, T(v) = T_0, A)}{f(T(v) = T_0, A)} = \frac{f(v, A)\,\delta(T(v) - T_0)}{f(T(v) = T_0, A)}. \tag{41}$$

Pentru obţinerea relaţiei (41), am avut în vedere faptul că $T(v)$ este dependentă funcţional de v, deci densitatea comună $f(v, T(v) = T_0, A)$ este nenulă doar atunci când v satisface $T(v) = T_0$. Densitatea comuna (numărătorul fracţiei) va fi deci $f(v, A)\,\delta(T(v) - T_0)$, δ fiind distribuţia Dirac. Deoarece $T(v) \sim \mathcal{N}\left(nA, n\sigma^2\right)$, avem:

$$\begin{aligned} f(v, A)\,\delta(T(v) - T_0) &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(x_i - A)^2\right)\delta(T(v) - T_0) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1} x_i^2 - 2AT(v) + nA^2\right)\right)\delta(T(v) - T_0) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1} x_i^2 - 2AT_0 + nA^2\right)\right)\delta(T(v) - T_0). \end{aligned}$$

Din (41),

$$f\left(v|T\left(v\right)=T_0,A\right) = \frac{\frac{1}{\left(2\pi\sigma^2\right)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}x_i^2\right)\exp\left(-\frac{1}{2\sigma^2}\left(-2AT_0+nA^2\right)\right)}{\frac{1}{\sqrt{2\pi n\sigma^2}}\exp\left(-\frac{1}{2n\sigma^2}\left(T_0-nA\right)^2\right)}\delta\left(T\left(v\right)-T_0\right)$$
$$= \frac{\sqrt{n}}{\left(2\pi\sigma^2\right)^{(n-1)/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}x_i^2\right)\exp\left(\frac{T_0^2}{2n\sigma^2}\right)\delta\left(T\left(v\right)-T_0\right),$$

cantitate care nu depinde de A. Concluzionăm că $\sum_{i=0}^{n-1}X_i$ este o statistică suficientă pentru estimarea parametrului A.

Acest exemplu indică o metodă pentru a verifica că o statistică este suficientă. Însă, în multe situaţii, problema determinării densităţii condiţionate este dificilă şi, prin urmare, este nevoie de o abordare mai uşoară. De asemenea, în general, o altă chestiune, chiar mai importantă, care trebuie avută în vedere, este *identificarea potenţialei statistici suficiente*. Un instrument important care ne permite rezolvarea acestei ultime chestiuni este reprezentat de *Teorema de factorizare Neyman-Fisher*.

Prezentăm, pentru început acest rezultat, după care, îl vom folosi pentru identificarea statisticilor suficiente în cadrul câtorva exemple reprezentative.

Teorema 2 (Teorema de factorizare Fisher-Neyman pentru parametru scalar) *Dacă putem factoriza densitatea de repartiţie*

$$f\left(v,\theta\right)=g\left(T\left(v,\theta\right)\right)h\left(v\right), \tag{42}$$

unde g este o funcţie dependentă de v doar prin intermediul lui $T\left(v\right)$, iar h este o funcţie dependentă doar de v,atunci $T\left(V\right)$ este o statistică suficientă pentru parametrul θ. Reciproc, dacă $T\left(V\right)$ este o statistică suficientă pentru θ, atunci densitatea de repartiţie a vectorului aleator de selecţie (adică funcţia de verosimilitate) poate fi factorizată precum în formula (42).

Demonstraţie. Considerăm densitatea de repartiţie comună (vectorială) $f\left(v,T\left(v\right),\theta\right)$. Pentru evaluarea acestei densităţi, menţionăm că $T\left(v\right)$ este dependentă funcţional de v. Prin urmare, densitatea comună va fi zero atunci când este calculată în $v=v_0$, $T\left(v\right)=T_0$ cu excepţia cazului când $T\left(v_0\right)=T_0$. Avem astfel

$$f\left(v,T\left(v\right)=T_0,\theta\right)=f\left(v,\theta\right)\delta\left(T\left(v\right)-T_0\right),$$

formulă ce va fi utilizată în cadrul demonstraţiei teoremei.

Un al doilea rezultat de care ne vom folosi constă în modul de a reprezenta densitatea de repartiţie a unei funcţii ce depinde de mai multe variabile aleatoare. Astfel, dacă $Y=g\left(X\right)$, unde X este un vector aleator $n-$dimensional, iar $g:\mathbb{R}^n\to\mathbb{R}$, atunci densitatea de repartiţie a variabilei aleatoare Y poate fi scrisă astfel:

$$f\left(y\right)=\int\cdots\int_{\mathbb{R}^n}f\left(v\right)\delta\left(y-g\left(v\right)\right)dv \tag{43}$$

Această reprezentare este echivalentă cu formula clasică a transformării variabilelor aleatoare. Dacă, de exemplu X este o variabilă aleatoare $1-$dimensională, atunci

$$\delta\left(y-g\left(x\right)\right)=\sum_{i=1}^{k}\frac{\delta\left(x-x_i\right)}{\left|g'\left(x\right)\right|_{x=x_i}},\quad \text{unde}\quad \left\{x_1,...,x_k\right\}\text{ sunt soluţiile ecuaţiei }y=g\left(x\right).$$

Dacă înlocuim această formulă în (43) obţinem formula clasică de transformare a variabilelor aleatoare.

Demonstrăm acum, pentru început, că $T\left(v\right)$ este o statistică suficientă, atunci când formula de factorizare are loc. Avem, datorită formulei de factorizare aplicată pentru deducerea ultimei egalităţi,

$$f\left(v|T\left(v\right)=T_0,\theta\right)=\frac{f\left(v,T\left(v\right)=T_0,\theta\right)}{f\left(T\left(v\right)=T_0,\theta\right)}=\frac{f\left(v,\theta\right)\delta\left(T\left(v\right)-T_0\right)}{f\left(T\left(v\right)=T_0,\theta\right)}=\frac{g\left(T\left(v\right)=T_0,\theta\right)h\left(v\right)\delta\left(T\left(v\right)-T_0\right)}{f\left(T\left(v\right)=T_0,\theta\right)} \tag{44}$$

Dar, în conformitate cu (43),

$$f\left(T\left(v\right)=T_0,\theta\right) = \int\cdots\int_{\mathbb{R}^n}f\left(v,\theta\right)\delta\left(T\left(v\right)-T_0\right)dv=\int\cdots\int_{\mathbb{R}^n}g\left(T\left(v\right)=T_0,\theta\right)h\left(v\right)\delta\left(T\left(v\right)-T_0\right)dv$$
$$= g\left(T\left(v\right)=T_0,\theta\right)\int\cdots\int_{\mathbb{R}^n}h\left(v\right)\delta\left(T\left(v\right)-T_0\right)dv.$$

Ultima egalitate din formula anterioară are loc deoarece integrala este zero pe complementara suprafeţei din $\mathbb{R}^n$ pentru care $T(v) = T_0$. Pe această suprafaţă, g este constantă. Inserăm formula obţinută în (44) şi rezultă:

$$f(v|T(v) = T_0, \theta) = \frac{h(v)\,\delta(T(v) - T_0)}{\int \cdots \int_{\mathbb{R}^n} h(v)\,\delta(T(v) - T_0)\,dv},$$

formulă ce nu depinde de parametrul θ. În concluzie, $T(V)$ este o statistică suficientă.

Pentru cea de a doua parte a teoremei, arătăm că, dacă $T(V)$ este o statistică suficientă, atunci formula de factorizare trebuie să aibă loc. Considerăm densitatea de repartiţie comună

$$f(v, T(v) = T_0, \theta) = f(v|T(v) = T_0, \theta)\, f(T(v) = T_0, \theta) \tag{45}$$

şi subliniem faptul că $f(v, T(v) = T_0, \theta) = f(v, \theta)\,\delta(T(v) - T_0)$. Deoarece $T(V)$ este o statistică suficientă, densitatea condiţionată nu trebuie să depindă de parametrul θ şi, prin urmare, o putem scrie ca $f(v|T(v) = T_0)$. Acum, deoarece avem $T(v) = T_0$ (ceea ce reprezintă o suprafaţă din $\mathbb{R}^n$), **densitatea condiţionată este nevidă doar pe această suprafaţă**. Notăm $f(v|T(v) = T_0) = w(v)\,\delta(T(v) - T_0)$, unde

$$\int \cdots \int_{\mathbb{R}^n} w(v)\,\delta(T(v) - T_0)\,dv = 1. \tag{46}$$

Obţinem, din (45),

$$f(v, \theta)\,\delta(T(v) - T_0) = w(v)\,\delta(T(v) - T_0)\, f(T(v) = T_0, \theta) \tag{47}$$

Definim acum

$$w(v) := \frac{h(v)}{\int \cdots \int_{\mathbb{R}^n} h(v)\,\delta(T(v) - T_0)\,dv},$$

astfel încât (46) să fie satisfăcută. Reprezentarea (47) devine:

$$f(v, \theta)\,\delta(T(v) - T_0) = \frac{h(v)\,\delta(T(v) - T_0)}{\int \cdots \int_{\mathbb{R}^n} h(v)\,\delta(T(v) - T_0)\,dv} f(T(v) = T_0, \theta),$$

sau, echivalent,

$$f(v, \theta) = g(T(v) = T_0, \theta)\, h(v), \quad \text{unde} \quad g(T(v) = T_0, \theta) := \frac{f(T(v) = T_0, \theta)}{\int \cdots \int_{\mathbb{R}^n} h(v)\,\delta(T(v) - T_0)\,dv}.$$

Scrisă sub această formă, densitatea de repartiţie a statisticii suficiente poate fi determinată cu ajutorul formulei de factorizare, deoarece

$$f(T(v) = T_0, \theta) = g(T(v) = T_0, \theta) \int \cdots \int_{\mathbb{R}^n} h(v)\,\delta(T(v) - T_0)\,dv.$$

Demonstraţia este, în acest moment, încheiată.

Trebuie menţionat faptul că, uneori, nu este evident faptul că densitatea de repartiţie poate fi factorizată sub forma dorită. În aceste situaţii, este posibil să nu existe o statistică suficientă. Prezentăm în continuare câteva exemple în care utilizăm instrumentul teoretic prezentat anterior.

Exemplul 2.2 ***Mărimea unui curent continuu aflat într-un WGN.*** *Reexaminăm problema discutată în exemplul precedent. Densitatea era dată de formula (40), unde σ^2 era presupus a fi cunoscută. Pentru a demonstra că există o factorizare, observăm că:*

$$\sum_{i=0}^{n-1} (x_i - A)^2 = \sum_{i=0}^{n-1} x_i^2 - 2A \sum_{i=0}^{n-1} x_i + nA^2,$$

ceea ce face ca densitatea să poată fi factorizată astfel:

$$f(v, A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(nA^2 - 2A\sum_{i=0}^{n-1} x_i\right)\right)}_{=:g(T(v),A)} \underbrace{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1} x_i^2\right)}_{=:h(v)}.$$

Este evident că $T(V)$ este o statistică suficientă pentru parametrul A. Observăm că $T'(V) = 2\sum_{i=0}^{n-1} X_i^2$ este, de asemenea, o statistică suficientă pentru A. Mai mult, se poate arăta că orice funcţie bijectivă ce are ca argument pe $\sum_{i=0}^{n-1} X_i$ este o statistică suficientă pentru A. Prin urmare, statisticile suficiente sunt unice până la o transformare bijectivă.

Exemplul 2.3 ***Puterea unui WGN***. *Considerăm acum densitatea dată de (40), cu $A = 0$ şi dispersia σ^2 fiind parametrul necunoscut. Avem astfel,*

$$f(v,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1} x_i^2\right) \cdot 1 = g\left(T\left(v\right), A\right) h\left(v\right).$$

Din Teorema de factorizare, se poate arăta că $T\left(V\right) := \sum_{i=0}^{n-1} X_i^2$ este o statistică suficientă pentru parametrul σ^2. Pentru aceasta, se arată că $f\left(v|\sum_{i=0}^{n-1} x_i^2 = T_0, \sigma^2\right)$ nu depinde de σ^2, plecând de la faptul că densitatea de repartiţie a variabilei aleatoare

$$S := \sum_{i=0}^{n-1} \frac{X_i^2}{\sigma^2} \sim \chi\left(n,1\right), \quad \textit{adică} \quad f_S\left(x\right) = \frac{1}{2^{n/2}\Gamma\left(n/2\right)} \exp\left(-s/2\right) s^{n/2-1} \mathbf{1}_{\{s\geq 0\}}.$$

Exemplul 2.4 ***Parametrul fazei unui curent sinusoidal***. *Reconsiderăm Exemplul 1.4, în care dorim să estimăm parametrul fazei unui curent de tip sinusoidal aflat într-un WGN:*

$$X_i = A\cos\left(2\pi f_0 i + \phi\right) + W_i, \quad i \in \{0,1,...,n-1\}.$$

Presupunem amplitudinea A şi frecvenţa f_0,precum şi dispersia σ^2 a zgomotului alb cunoscute. Densitatea comună (funcţia de verosimilitate) este:

$$\begin{aligned}
f\left(v,\phi\right) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}\left(x_i - A\cos\left(2\pi f_0 i + \phi\right)\right)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1} x_i^2 - 2A\sum_{i=0}^{n-1} x_i\cos\left(2\pi f_0 i + \phi\right) + \sum_{i=0}^{n-1} A^2\cos^2\left(2\pi f_0 i + \phi\right)\right)\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1} x_i^2 - 2A\left(\sum_{i=0}^{n-1} x_i\cos\left(2\pi f_0 i\right)\right)\cos\phi + 2A\left(\sum_{i=0}^{n-1} x_i\sin\left(2\pi f_0 i\right)\right)\sin\phi\right.\right. \\
&\quad \left.\left. + \sum_{i=0}^{n-1} A^2\cos^2\left(2\pi f_0 i + \phi\right)\right)\right).
\end{aligned}$$

În acest exemplu nu apare faptul că densitatea este factorizabilă aşa cum este nevoie în Teorema Neyman-Fisher, adică nu ar exista nicio statistică suficientă pentru parametrul fazei, ϕ. Cu toate acestea, putem scrie sub forma factorizată următoare:

$$\begin{aligned}
&f\left(v,\phi\right) \\
&= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1} A^2\cos^2\left(2\pi f_0 i + \phi\right) - 2AT_1\left(x\right)\cos\phi + 2AT_2\left(x\right)\sin\phi\right)\right)}_{=:g(T_1(v),T_2(v),\phi)} \underbrace{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1} x_i^2\right)}_{=:h(v)},
\end{aligned}$$

unde

$$T_1\left(v\right) := \sum_{i=0}^{n-1} x_i\cos\left(2\pi f_0 i\right) \qquad \textit{şi} \qquad T_2\left(v\right) := \sum_{i=0}^{n-1} x_i\sin\left(2\pi f_0 i\right).$$

O generalizare a Teoremei Neyman-Fisher afirmă că $\left(T_1\left(V\right), T_2\left(V\right)\right)$ sunt, ca pereche, statistici suficiente pentru parametrul ϕ, fără a exista însă o statistică suficientă singulară. Motivul pentru care ne vom restrange atenţia doar asupra statisticilor suficiente singulare va fi clarificat în secţiunea următoare.

Conceptul de statistici suficiente grupate trebuie privit astfel. Spunem că r statistici $T_1\left(V\right), T_2\left(V\right), ..., T_r\left(V\right)$ formează o statistică suficientă grupată dacă densitatea condiţionată $f\left(v, T_1\left(v\right), T_2\left(v\right), ..., T_r\left(v\right), \theta\right)$ nu depinde de θ. Generalizarea Teoremei Neyman-Fisher spune că, dacă densitatea comună (a vectorului aleator de selecţie) $f\left(v,\theta\right)$poate fi factorizată sub forma (cititorul interesat poate consulta Kendall, Stuart [4]):

$$f\left(v,\theta\right) = g\left(T_1\left(v\right), T_2\left(v\right), ..., T_r\left(v\right), \theta\right) h\left(v\right), \tag{48}$$

atunci $r-$uplul $\{T_1\left(V\right), T_2\left(V\right), ..., T_r\left(V\right)\}$ formează o statistică suficientă grupată pentru parametrul θ.

Observaţia 2.1 *Este evident faptul că datele observate furnizează întotdeauna o astfel de statistică grupată deoarece, dacă stabilim $r = n$ şi $T_{i+1}\left(V\right) := X_i$, $i = 0,1,...,n-1$, atunci putem alege $g = f\left(v,\theta\right)$ şi $h = 1$, iar condiţia (48) va fi identic satisfăcută. Desigur, foarte rar, ele formează mulţimea minimală de statistici suficiente.*

2.3 Utilizarea suficienţei în determinarea estimatorilor MVU

Presupunând că am reuşit să determinăm o statistică suficientă $T(V)$ pentru parametrul θ, putem utiliza apoi *Teorema Rao-Blackwell-Lehmann-Scheffe (Teorema RBLS)* pentru a găsi un estimator *MVU*. Prezentăm, pentru început, un exemplu care să ilustreze abordarea problemei, după care vom introduce cadrul teoretic de lucru.

Exemplul 2.5 ***Mărimea curentului continuu în WGN.*** *Vom continua analiza Exemplului 2.2. Deşi deja am obţinut faptul că $\hat{A} = \bar{X}$ este un estimator MVU (deoarece este eficient), utilizăm acum Teorema RBLS, care poate fi folosită chiar şi atunci când nu există un estimator eficient, deci metoda CRLB nu este viabilă. Procedura pentru determinarea estimatorului MVU, $\hat{A}$, poate fi aplicată în două moduri. Ambele se bazează pe statistica suficientă $T(V) = \sum_{i=0}^{n-1} X_i$:*

1. *Găsim orice estimator nedeplasat al parametrului A, să presupunem $\bar{A} = X_0$, şi determinăm apoi $\hat{A} = \mathbb{E}\left(\bar{A}|T\right)$.*
2. *Determinăm o funcţie g astfel încât $\hat{A} = g(T)$ este un estimator nedeplasat pentru parametrul A.*

Pentru prima abordare putem pleca de la estimatorul nedeplasat $\bar{A} = X_0$ şi determinăm $\hat{A} = \mathbb{E}(X_0|\sum_{i=0}^{n-1} X_i)$. Pentru a realiza aceasta avem nevoie de câteva proprietăţi ale densităţii condiţionate Gaussiene. Astfel, pentru un vector aleator Gaussian $(X,Y)^t$, cu media $\mu = (\mathbb{E}(X), \mathbb{E}(Y))^t$ şi matricea de covarianţă

$$C = \begin{pmatrix} D^2(X) & cov(X,Y) \\ cov(Y,X) & D^2(Y) \end{pmatrix},$$

densitatea de repartiţie este

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi \det^{1/2}(C)} \exp\left(-\frac{1}{2}\begin{pmatrix} x-\mathbb{E}(X) \\ y-\mathbb{E}(Y) \end{pmatrix}^t C^{-1} \begin{pmatrix} x-\mathbb{E}(X) \\ y-\mathbb{E}(Y) \end{pmatrix}\right).$$

Densităţile condiţionate $f(x|y)$ şi $f(y|x)$ sunt, de asemenea, Gaussiene şi au loc următoarele formule pentru media şi dispersia condiţionată (demonstraţiile pot fi studiate în Kay [2, Theorem 10.1 şi Theorem 10.2]):

$$\begin{cases} \mathbb{E}(X|Y) &= \displaystyle\int_{-\infty}^{+\infty} x f(x|y)\,dx = \int_{-\infty}^{+\infty} x \frac{f_{(X,Y)}(x,y)}{f_Y(y)} dx = \mathbb{E}(X) + \frac{cov(X,Y)}{D^2(Y)}(Y - \mathbb{E}(Y)). \\ D^2(X|Y) &= D^2(X) - \dfrac{cov^2(X,Y)}{D^2(Y)}. \end{cases} \tag{49}$$

Aplicăm acum aceste formule pentru variabilele aleatoare $X = X_0$ şi $Y = \sum_{i=0}^{n-1} X_i$. Observăm, de asemenea, că:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X_0 \\ \sum_{i=0}^{n-1} X_i \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}}_{=:L} \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_{n-1} \end{pmatrix}.$$

Prin urmare,

$$(X,Y) \sim \mathcal{N}(\mu, C), \qquad \textit{unde}$$

$$\mu = L\mathbb{E}(X) = LA\mathbf{1} = \begin{pmatrix} A \\ nA \end{pmatrix} \qquad \textit{şi} \qquad C = \sigma^2 LL^t = \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & n \end{pmatrix},$$

deoarece reprezintă o transformare liniară a unui vector aleator Gausian. Formula (49) conduce la

$$\hat{A} = \mathbb{E}(X|Y) = A + \frac{\sigma^2}{n\sigma^2}\left(\sum_{i=0}^{n-1} X_i - nA\right) = \frac{1}{n}\sum_{i=0}^{n-1} X_i,$$

care este un estimator MVU. Această abordare, care presupune evaluarea mediei condiţionate, este, de regulă, greu abordabilă din punct de vedere matematic.

În ceea ce priveşte cea de a doua abordare, trebuie să găsim o funcţie g astfel încât

$$\hat{A} = g\left(\sum_{i=0}^{n-1} X_i\right)$$

să fie un estimator nedeplasat pentru A. Dacă considerăm $g(x) = x/n$, obţinem imediat un estimator MVU. Această metodă este mai uşor de aplicat şi este mai des utilizată în practică.

Teorema 3 (Rao-Blackwell-Lehmann-Scheffe, cazul scalar) *Dacă $\bar{\theta}$ este un estimator nedeplasat pentru parametrul θ şi $T(V)$ este o statistică suficientă pentru θ, atunci $\hat{\theta} = \mathbb{E}\left(\bar{\theta}|T(V)\right)$ este:*

1. *un estimator valid pentru θ (adică nu depinde de θ);*
2. *nedeplasat;*
3. *de dispersie mai mică sau egală cu dispersia lui $\bar{\theta}$, pentru toţi θ.*

În plus, dacă statistica suficientă este completă, atunci $\hat{\theta}$ este estimatorul MVU.

Demonstraţie. Vom analiza cazul parametrului $1-$dimensional, θ. Pentru a demonstra (1), observăm că:

$$\hat{\theta} = \mathbb{E}\left(\bar{\theta}|T(V)\right) = \int \cdots \int_{\mathbb{R}^n} \bar{\theta}(v) f(v|T(v), \theta)\, dv \tag{50}$$

Din definiţia unei statistici suficiente, $f(v|T(v), \theta)$ nu depinde de θ, ci de v şi de $T(v)$. Prin urmare, în urma integrării, rezultatul va fi o funcţie de T şi nu va depinde de θ.

Pentru punctul (2), adică $\hat{\theta}$ este nedeplasat, din (50) şi din faptul că $\hat{\theta}$ este dependent doar de T, avem:

$$\begin{aligned}\mathbb{E}(\hat{\theta}) &= \int \cdots \int_{\mathbb{R}^{n+1}} \bar{\theta}(v) f(v|T(v), \theta)\, dv\, f(T(v), \theta)\, dT = \int \cdots \int_{\mathbb{R}^n} \bar{\theta}(v) \int_{-\infty}^{+\infty} f(v|T(v), \theta) f(T(v), \theta)\, dT dv \\ &= \int \cdots \int_{\mathbb{R}^n} \bar{\theta}(v) f(v, \theta)\, dv = \mathbb{E}\left(\bar{\theta}\right) = \theta \quad \text{deoarece } \bar{\theta} \text{ este estimator nedeplasat.}\end{aligned}$$

Pentru a demonstra punctul (3), şi anume că $D^2(\hat{\theta}) \leq D^2\left(\bar{\theta}\right)$, procedăm astfel:

$$D^2\left(\bar{\theta}\right) = \mathbb{E}\left(\left(\bar{\theta} - \mathbb{E}\left(\bar{\theta}\right)\right)^2\right) = \mathbb{E}\left(\left(\bar{\theta} - \hat{\theta} + \hat{\theta} - \theta\right)^2\right) = \mathbb{E}\left(\left(\bar{\theta} - \hat{\theta}\right)^2\right) + 2\mathbb{E}\left(\left(\bar{\theta} - \hat{\theta}\right)\left(\hat{\theta} - \theta\right)\right) + D^2(\hat{\theta}).$$

Ştim că $\hat{\theta}$ este o funcţie ce depinde doar de T. De aici rezultă că:

$$\mathbb{E}_{T,\bar{\theta}}\left(\left(\bar{\theta} - \hat{\theta}\right)\left(\hat{\theta} - \theta\right)\right) = \mathbb{E}_T \mathbb{E}_{\bar{\theta}|T}\left(\left(\bar{\theta} - \hat{\theta}\right)\left(\hat{\theta} - \theta\right)\right) = \mathbb{E}_T \mathbb{E}_{\bar{\theta}|T}\left(\bar{\theta} - \hat{\theta}\right)\left(\hat{\theta} - \theta\right) = \mathbb{E}_T \underbrace{\left(\mathbb{E}_{\bar{\theta}|T}\left(\bar{\theta}|T\right) - \hat{\theta}\right)}_{=\hat{\theta}-\hat{\theta}=0}\left(\hat{\theta} - \theta\right) = 0.$$

Am obţinut astfel rezultatul dorit, adică

$$D^2(\bar{\theta}) = \mathbb{E}((\bar{\theta} - \hat{\theta})^2) + D^2(\hat{\theta}) \geq D^2(\hat{\theta}).$$

O statistică este completă dacă există o singură funcţie ce depinde de statistică şi care este nedeplasată. Justificăm acum că $\hat{\theta} = \mathbb{E}\left(\bar{\theta}|T(V)\right)$ este estimatorul *MVU*. Considerăm toţi estimatorii nedeplasaţi ai lui θ. Determinând $\mathbb{E}\left(\bar{\theta}|T(V)\right)$ putem micşora dispersia estimatorului (punctul (3)) şi rămânem în clasa estimatorilor admisibili (punctul (2)). Dar $\mathbb{E}\left(\bar{\theta}|T(V)\right)$ este o funcţie ce depinde doar de statistica suficientă $T(V)$ deoarece

$$\hat{\theta} = \mathbb{E}\left(\bar{\theta}|T(V)\right) = \int \cdots \int_{\mathbb{R}^n} \bar{\theta}(v) f\left(\bar{\theta}|T(v)\right) d\bar{\theta} = g(T(v)) \tag{51}$$

Dacă $T(V)$ este complet, există doar o singură funcţie de T ce este estimator nedeplasat. Prin urmare, $\hat{\theta}$ este unic, independent de $\bar{\theta}$ ales din clasa estimatorilor nedeplasaţi. Fiecare $\bar{\theta}$ va furniza acelaşi estimator $\hat{\theta}$. Cum dispersia lui $\hat{\theta}$ este mai mică decât dispersia oricărui estimator $\bar{\theta}$ din clasă (punctul (3)), rezultă că $\hat{\theta}$ este estimatorul *MVU*. Sumarizând, estimatorul *MVU* se poate obţine considerând orice estimator nedeplasat şi aplicându-i transformarea dată de (51). Alternativ, cum există o unică funcţie dependentă de statistica suficientă ce duce la un estimator nedeplasat, este suficient să o identificăm. Demonstraţia este încheiată.

Pentru ultima afirmaţie din Teorema 3, am văzut în Exemplul 2.5 că $g(\sum_{i=0}^{n-1} X_i) = \sum_{i=0}^{n-1} X_i/n$. Proprietatea de completitudine depinde de densitatea de repartiţie a vectorului de selecţie V, ce determină densitatea statisticii suficiente. În general, este dificil de arătat că o statistică suficientă este completă, un studiu detaliat al acestei problematici putând fi analizat în Kendall, Stuart [4]. Vom prezenta în continuare două exemple, pentru a familiariza cititorul cu conceptul de completitudine al statisticilor suficiente.

Exemplul 2.6 ***Completitudinea statisticii suficiente.*** *Pentru estimarea parametrului A, statistica suficientă $T = \sum_{i=0}^{n-1} X_i$ este completă, în sensul că există o unică funcţie g pentru care $\mathbb{E}(g(T)) = A$. Presupunem, prin reducere la absurd, că ar mai exista încă o funcţie h, cu aceeaşi proprietate: $\mathbb{E}(h(T)) = A$. Rezultă deci că*

$$\mathbb{E}\left(g(T) - h(T)\right) = A - A = 0, \quad \textit{pentru toţi } A.$$

Deoarece $T \sim \mathcal{N}\left(nA, n\sigma^2\right)$, eglitatea de mai sus înseamnă:

$$\int_{-\infty}^{+\infty} u\left(T\right) \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2}\left(T - nA\right)^2\right) dT = 0, \quad \textit{pentru toţi } A,$$

unde am notat $u\left(T\right) = g\left(T\right) - h\left(T\right)$. Dacă notăm $\tau = T/n$ şi $u\prime\left(\tau\right) = u\left(n\tau\right)$, avem:

$$\int_{-\infty}^{+\infty} u\prime\left(\tau\right) \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{n}{2\sigma^2}\left(A - \tau\right)^2\right) d\tau = 0, \quad \textit{pentru toţi } A, \tag{52}$$

formulă care reprezintă convoluţia funcţiei $u\prime\left(\tau\right)$ cu un puls Gaussian, $w\left(\tau\right)$. Cum egalitatea (52) are loc pentru toţi A, $u\prime\left(\tau\right)$ trebuie să fie identic zero. De aici rezultă că $0 = u\left(n\tau\right) = g\left(n\tau\right) - h\left(n\tau\right)$, pentru orice τ, adică $g \equiv h$.

Exemplul 2.7 ***Lipsa de completitudine a statisticii suficiente.*** *Considerăm estimatorul lui A dat de:*

$$X_0 = A + W_0, \quad \textit{unde} \quad W_0 \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right).$$

O statistică suficientă este X_0, furnizând unica dată disponibilă şi, prin urmare, X_0 este un estimator nedeplasat pentru A. Am putea concluziona că $g\left(X_0\right) = X_0$ este un candidat viabil pentru estimatorul MVU. Ar trebui verificat că este o statistică suficientă completă. Similar exemplului anterior, presupunem că există încă o funcţie h ce verifică $\mathbb{E}\left(h\left(X_0\right)\right) = A$ şi încercăm să arătăm că $h \equiv g$. Notăm iar $u\left(T\right) = g\left(T\right) - h\left(T\right)$ şi căutăm posibilele soluţii v pentru ecuaţia:

$$\int_{-\infty}^{+\infty} u\left(T\right) f\left(v, A\right) dv = 0, \quad \textit{pentru toţi } A.$$

Pentru acestă problemă, $V = X_0 = T$ şi rezultă

$$\int_{-\infty}^{+\infty} u\left(T\right) f\left(T, A\right) dT = 0 \quad \Longleftrightarrow \quad \int_{A-1/2}^{A+1/2} u\left(T\right) dT = 0, \qquad \textit{pentru toţi } A,$$

deoarece

$$f\left(T, A\right) = \begin{cases} 1, & A - 1/2 \leq T \leq A + 1/2 \\ 0, & \textit{în caz contrar.} \end{cases}$$

O soluţie a ecuaţiei de mai sus este $u\left(T\right) = \sin\left(2\pi T\right)$, iar de aici rezultă că $\sin\left(2\pi T\right) = g\left(T\right) - h\left(T\right)$, sau, echivalent, $h\left(T\right) = g\left(T\right) - \sin\left(2\pi T\right)$. Obţinem astfel că estimatorul

$$\hat{A} = X_0 - 2\sin\left(2\pi X_0\right)$$

este, de asemenea, bazat pe statistica suficientă şi este un estimator nedeplasat pentru parametrul A. Cum am găsit cel puţin încă un estimator nedeplasat ce e funcţie de statistica suficientă, putem concluziona că statistica suficientă nu este completă. Teorema RBLS nu mai poate fi aplicată şi nu putem trage concluzia că $\hat{A} = X_0$ este estimatorul MVU.

Pentru a concluziona, prezentă etapele ce trebuie parcurse pentru determinarea unui estimator *MVU*:

1. Determinăm o statistică suficientă pentru parametrul θ, notată cu $T\left(V\right)$, folosind Teorema de factorizare Neman-Fisher.

2. Determinăm dacă statistica suficientă este completă, iar dacă da, putem continua. În caz contrar, această abordare nu poate fi utilizată.

3. Determinăm o funcţie g, ce are ca argument statistica suficientă şi care produce un estimator nedeplasat, $\hat{\theta} = g\left(T\left(V\right)\right)$. Estimatorul *MVU* va fi $\hat{\theta}$. O variantă alternativă a acestui pas o constituie calcularea $\hat{\theta} = \mathbb{E}(\bar{\theta}|T\left(V\right))$, unde $\bar{\theta}$ este un estimator nedeplasat oarecare. Această ultimă abordare este, de regulă, greu de aplicat în practică, tocmai datorită modului de evaluare a mediei condiţionate.

Exemplul 2.8 ***Valoarea medie a zgomotului uniform.*** *Presupunem că avem un model în care datele observate x_i sunt corespunzătoare variabilelor de selecţie $X_i = W_i$, unde W_i, $i \in \{0, 1, ..., n-1\}$, sunt zgomote independente stochastic, identic repartizate $\mathcal{U}(0, \beta)$, $\beta > 0$. Ne propunem să determinăm un estimator MVU pentru media $\theta = \beta/2$.*

Abordarea prin intermediul CRLB pentru determinarea unui estimator eficient şi, prin urmare, care să fie şi un estimator MVU nu poate fi utilizată pentru această situaţie deoarece densitatea de repartiţie a vectorului aleator de selecţie, adică funcţia de verosimilitate, nu satisface condiţiile de regularitate, având:

$$\mathbb{E}\left(\frac{\partial \ln \mathcal{L}(v, \theta)}{\partial \theta}\right) \neq 0, \quad \textit{pentru toţi } \theta.$$

Un estimator natural pentru θ este media de selecţie $\bar{X} = (\sum_{i=0}^{n-1} X_i)/n$. Acesta este nedeplasat şi are dispersia $D^2(\bar{X}) = \beta^2/(12n)$. Pentru a determina dacă acest estimator este MVU vom urma paşii descrişi înaintea acestui exemplu. Construim, pentru început, funcţia în scară

$$u(x) = \mathbf{1}_{\{x \geq 0\}} = \begin{cases} 1, & \textit{dacă } x \geq 0, \\ 0, & \textit{dacă } x < 0. \end{cases}$$

Obţinem astfel densitatea fiecărei variabile aleatoare de selecţie, precum şi pentru funcţia de verosimilitate:

$$\begin{cases} f(x_i, \theta) &= \dfrac{1}{\beta}\left(u(x_i) - u(x_i - \beta)\right), \quad \textit{unde} \quad \beta = 2\theta, \\ \mathcal{L}(v, \theta) &= \dfrac{1}{\beta^n}\displaystyle\prod_{i=0}^{n-1}\left(u(x_i) - u(x_i - \beta)\right) \neq 0 \quad \textit{doar dacă } 0 < x_i < \beta, \textit{ pentru toţi indicii } i. \end{cases}$$

Astfel, obţinem următoarea factorizare a densităţii vectoriale:

$$\mathcal{L}(v, \theta) = \mathcal{L}(x_0, ..., x_{n-1}, \theta) = \mathbf{1}_{\{0 < x_i < \beta,\, \forall i\}} = \underbrace{\beta^{-n} u\left(\beta - \max_{i=\overline{0,n-1}} x_i\right)}_{=:g(T(v),\theta)} \underbrace{u\left(\min_{i=\overline{0,n-1}} x_i\right)}_{=:h(v)}.$$

Apelăm la Teorema de factorizare Neyman-Fisher şi obţinem că ultima statistică de ordine $T(V) = \max_{i=\overline{0,n-1}} X_i$ este o statistică suficientă pentru parametrul θ.

În următoarea etapă determinăm o funcţie ce are ca argument statistica suficientă şi care produce o statistică nedeplasată. Calculăm, pentru început, repartiţia statisticii de ordine T :

$$F_T(\xi) = \mathbb{P}(T \leq \xi) = \mathbb{P}(X_0 \leq \xi, ..., X_{n-1} \leq \xi) = \prod_{i=0}^{n-1} \mathbb{P}(X_i \leq \xi) = (\mathbb{P}(X_0 \leq \xi))^n.$$

Densitatea de repartiţie a lui T se obţine prin derivarea funcţiei de repartiţie calculate anterior:

$$\begin{aligned} f_T(\xi) &= n(\mathbb{P}(X_0 \leq \xi))^{n-1} \frac{d\mathbb{P}(X_0 \leq \xi)}{d\xi} = n(\mathbb{P}(X_0 \leq \xi))^{n-1} f_{X_0}(\xi) = n(\mathbb{P}(X_0 \leq \xi))^{n-1} \frac{1}{\beta}\mathbf{1}_{\{0<\xi<\beta\}}. \\ &= n\left(\frac{\xi}{\beta}\right)^{n-1} \frac{1}{\beta}\mathbf{1}_{\{0<\xi<\beta\}}. \end{aligned}$$

Media statisticii suficiente T este:

$$\mathbb{E}(T) = \int_{-\infty}^{+\infty} \xi f_T(\xi)\, d\xi = \int_0^{\beta} \xi n \left(\frac{\xi}{\beta}\right)^{n-1} \frac{1}{\beta} d\xi = \frac{n}{n+1}\beta = \frac{2n}{n+1}\theta.$$

Pentru a face acest estimator nedeplasat, trebuie ca $2n/(n+1) = 1$ şi vom considera estimatorul $\hat{\theta} := ((n+1)/(2n))\, T$, ceea ce spune că estimatorul MVU este

$$\hat{\theta} = \frac{n+1}{2n} \max_{i=\overline{0,n-1}} X_i.$$

Cumva, contrar intuiţiei, ***media de selecţie nu este estimatorul MVU*** *pentru media zgomotului distribuit uniform. Dacă dorim să comparăm dispersia minimală cu cea a mediei de selecţie, observăm că:*

$$D^2(\hat{\theta}) = \left(\frac{n+1}{2n}\right)^2 D^2(T) = \left(\frac{n+1}{2n}\right)^2 \left(\int_0^{\beta} \xi^2 \frac{n\xi^{n-1}}{\beta^n} d\xi - \left(\frac{n\beta}{n+1}\right)^2\right) = \left(\frac{n+1}{2n}\right)^2 \frac{n\beta^2}{(n+1)^2(n+2)} = \frac{\beta^2}{4n(n+2)}.$$

Se observă că, pentru $n \geq 2$,

$$D^2(\bar{X}) = \frac{\beta^2}{12n} > \frac{\beta^2}{4n\left(n+2\right)} = D^2(\hat{\theta}).$$

Succesul acestui exemplu se bazează pe posibilitatea de a determina o statistică suficientă singulară. Dacă ne referim la Exemplul 2.4, a estimării parametrului fazei, acolo am avut nevoie de două statistici $T_1\left(V\right)$ *şi* $T_2\left(V\right)$*, fapt ce face abordarea prezentată dificil de aplicat.*

2.4 Extinderea la cazul parametrilor vectoriali

Extindem studiile precedente şi intenţionăm să determinăm un estimator vectorial *MVU*, de tipul $p \times 1$, pentru un parametru vectorial de aceeaşi dimensiune. Fiecare componentă a estimatorului trebuie să fie nedeplasată raportată la parametrul estimat şi să aibă dispersie (varianţă) minimă. În cazul vectorial, am putea avea mai multe statistici suficiente decât numărul de parametri estimaţi $(r > p)$, exact tot atâtea $(r = p)$, sau chiar mai puţine $(r < p)$. În exemplele abordate în această secţiune ne vom referi doar la primele două situaţii. O situaţie de dorit ar fi ca $r = p$, deoarece aceasta ar permite determinarea estimatorilor *MVU* prin transformarea statisticilor suficiente, păstrând proprietatea de nedeplasare.

O *statistică vectorială* $\mathbf{T}\left(V\right) = \left(T_1\left(V\right), T_2\left(V\right), ..., T_r\left(V\right)\right)^t$ se numeşte *suficientă* pentru estimarea parametrului vectorial θ dacă densitatea condiţionată a vectorului de selecţie în raport cu statistica, adică $f\left(V|\mathbf{T}\left(V\right), \theta\right)$, nu depinde de parametrul θ. Mai precis, aceasta înseamnă că $f\left(V|\mathbf{T}\left(V\right), \theta\right) = f\left(V|\mathbf{T}\left(V\right)\right)$. Cum, în general, pot exista mai multe astfel de statistici $\mathbf{T}\left(V\right)$, suntem interesaţi de *statistica suficientă minimală*, sau, altfel spus, de *statistica* $\mathbf{T}$ *de dimensiune minimală*. Pentru aceasta, facem apel tot la Teorema de factorizare Neyman-Fisher, în varianta sa vectorială.

Teorema 4 (Teorema de factorizare Fisher-Neyman pentru parametru vectorial) *Dacă putem factoriza densitatea de repartiţie a vectorului aleator de selecţie sub forma:*

$$\mathcal{L}\left(v, \theta\right) = g\left(\mathbf{T}\left(v\right), \theta\right) h\left(v\right), \tag{53}$$

unde g *este o funcţie cu valori în* $\mathcal{M}_{r\times 1}$*, ce depinde doar de* θ *şi de* v *prin intermediul lui* $\mathbf{T}\left(v\right)$ *şi* h *este o funcţie ce depinde doar de* v*, atunci* $\mathbf{T}\left(V\right)$ *este este o statistică suficientă pentru* θ*. Reciproc, dacă* $\mathbf{T}\left(V\right)$ *este este o statistică suficientă pentru* θ*, funcţia de verosimilitate poate fi factorizată precum în formula (53).*

Demonstraţia, pe care o vom omite aici, poate fi studiată în Kendall, Stuart [4]. Precizăm că, întotdeauna, va exista o mulţime de statistici suficiente, mulţimea de date oferite de V fiind, chiar ele, suficiente. Problema care trebuie cercetată este dimensiunea statisticii suficiente, precum şi concluzia oferită de Teorema 4.

Exemplul 2.9 ***Estimarea ansamblului parametrilor unui curent sinusoidal.*** *Presupunem că avem un curent de tip sinusoidal, aflat într-un WGN, pentru care datele observate sunt date de variabilele aleatoare de selecţie:*

$$X_i = A\cos\left(2\pi f_0 i\right) + W_i, \quad i \in \{0, 1, ..., n-1\},$$

unde amplitudinea A*, frecvenţa* f_0 *şi dispersia zgomotului,* σ^2*, sunt parametri necunoscuţi. Vectorul parametrilor va fi* $\theta = (A, f_0, \sigma^2)^t$*, iar densitatea comună a datelor este:*

$$\begin{aligned}
\mathcal{L}\left(v, \theta\right) &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}\left(x_i - A\cos\left(2\pi f_0 i\right)\right)^2\right) \\
&= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1}x_i^2 - 2A\sum_{i=0}^{n-1}x_i\cos\left(2\pi f_0 i\right) + \sum_{i=0}^{n-1}A^2\cos^2\left(2\pi f_0 i\right)\right)\right).
\end{aligned}$$

Datorită termenului $\sum_{i=0}^{n-1}x_i\cos\left(2\pi f_0 i\right)$*, cu* f_0 *necunoscut, nu putem reduce densitatea anterioară la forma dată de (53). Dacă, în schimb, frecvenţa ar fi cunoscută, vectorul parametrilor necunoscuţi este* $\theta = (A, \sigma^2)^t$*, iar densitatea se poate reprezenta imediat astfel:*

$$\mathcal{L}\left(v, \theta\right) = \underbrace{\frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1}x_i^2 - 2A\sum_{i=0}^{n-1}x_i\cos\left(2\pi f_0 i\right) - +\sum_{i=0}^{n-1}A^2\cos^2\left(2\pi f_0 i\right)\right)\right)}_{=:g(T(v),\theta)} \underbrace{1}_{=:h(v)},$$

unde

$$\mathbf{T}\left(v\right) := \left(\begin{array}{cc} \sum_{i=0}^{n-1}x_i\cos\left(2\pi f_0 i\right) & \sum_{i=0}^{n-1}x_i^2 \end{array}\right)^t.$$

Statistica $T(V)$ este o statistică suficientă pentru A şi σ^2, dar doar dacă f_0 este cunoscut.

Presupunem acum că avem:

$$X_i = A + W_i, \quad \text{unde } W_i \text{ este WGN}, \quad \text{pentru } i \in \{0, 1, ..., n-1\}$$

şi presupunem că amplitudinea A şi dispersia σ^2 a zgomotului alb Gaussian sunt parametri necunoscuţi. Putem folosi rezultatele obţinute în prima parte a exemplului pentru a determina o statistică suficientă de dimensiune 2 pentru estimarea mărimii curentului continuu şi a dispersiei zgomotului. Parametrul necunoscut este $\theta = (A, \sigma^2)^t$, iar statistica suficientă este:

$$T(V) = \left(\sum_{i=0}^{n-1} X_i \qquad \sum_{i=0}^{n-1} X_i^2 \right)^t. \tag{54}$$

Am văzut că, dacă σ^2 este cunoscut, $\sum_{i=0}^{n-1} X_i$ este o statistică suficientă pentru amplitudinea A (vezi Exemplul 2.2), iar dacă A este cunoscut ($A = 0$), $\sum_{i=0}^{n-1} X_i^2$ este o statistică suficientă pentru σ^2 (vezi Exemplul 2.3). În analiza de faţă, aceleaşi statistici sunt, în ansamblu, ca pereche, suficiente. Acest lucru nu este, întotdeauna, adevărat.

Cum, în exemplul precedent, avem doi parametri şi am obţinut două statistici suficiente, putem acum determina un estimator *MVU* pentru parametrul vectorial. Pentru a realiza aceasta, enunţăm varianta vectorială a Teoremei Rao-Blackwell-Lehmann-Scheffe.

Teorema 5 (Rao-Blackwell-Lehmann-Scheffe, cazul vectorial) *Dacă $\bar{\theta}$ este un estimator nedeplasat pentru parametrul θ şi $\mathbf{T}(V)$ este o statistică suficientă, $r \times 1$ dimensională, pentru θ, atunci $\hat{\theta} = \mathbb{E}\left(\bar{\theta} | \mathbf{T}(V)\right)$ este:*

1. *un estimator valid pentru θ (adică nu depinde de θ);*
2. *nedeplasat;*
3. *de dispersie mai mică sau egală cu dispersia lui $\bar{\theta}$ (compararea făcându-se componentă cu componentă), pentru toţi θ.*

În plus, dacă statistica suficientă este completă, atunci $\hat{\theta}$ este estimatorul MVU.

În cazul vectorial, completitudinea înseamnă că pentru o funcţie, $r \times 1-$dimensională, de $\mathbf{T}$, $\mathbf{u}(\mathbf{T})$, dacă

$$\mathbb{E}(\mathbf{u}(\mathbf{T})) = \int \cdots \int_{\mathbb{R}^r} \mathbf{u}(\mathbf{T}) \mathcal{L}(\mathbf{T}, \theta) \, d\mathbf{T} = \mathbf{0}, \quad \text{pentru toţi } \theta, \tag{55}$$

atunci $\mathbf{u}(\mathbf{T}) = \mathbf{0} \in \mathcal{M}_{r \times 1}(\mathbb{R})$, pentru toţi $\mathbf{T}$. Similar cazului $1-$dimensional pentru parametru, condiţia este greu verificabil în practică. Pentru a determina estimatorul *MVU* direct cu $\mathbf{T}(V)$, fără a mai evalua $\mathbb{E}\left(\bar{\theta} | \mathbf{T}(V)\right)$, dimensiunea statisticii suficiente ar trebui să fie egală cu dimensiunea vectorului parametrilor necunoscuţi, $r = p$. Dacă această condiţie este verificată, ar trebui doar să găsim o funcţie $p-$dimensională $\mathbf{g}$, astfel încât

$$\mathbb{E}(\mathbf{g}(\mathbf{T})) = \theta,$$

în ipoteza că $\mathbf{T}$ este o statistică suficientă completă.

Exemplul 2.10 (continuarea Exemplului 2.9) *Pentru statistica suficientă (54), media sa este:*

$$\mathbb{E}(\mathbf{T}(V)) = \begin{pmatrix} \mathbb{E}(T_1(V)) & \mathbb{E}(T_2(V)) \end{pmatrix} = \begin{pmatrix} nA & n\mathbb{E}(X_0^2) \end{pmatrix}^t = \begin{pmatrix} nA & n(\sigma^2 + A^2) \end{pmatrix}^t.$$

Desigur, am fi putut rezolva problema deplasării pentru prima componentă a lui $\mathbf{T}(V)$, prin împărţirea sa la n, dar ar fi dăunat celei de a doua componente. $T_2(V)$ estimează momentul teoretic de ordinul doi şi nu dispersia dorită. Dacă transformăm pe $\mathbf{T}(V)$ astfel:

$$\mathbf{g}(\mathbf{T}(V)) = \begin{pmatrix} \frac{1}{n} T_1(V) \\ \frac{1}{n} T_2(V) - \left(\frac{1}{n} T_1(V)\right)^2 \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum_{i=0}^{n-1} X_i^2 - \bar{X}^2 \end{pmatrix},$$

atunci, deoarece $\bar{X} \sim \mathcal{N}\left(A, \sigma^2/n\right)$,

$$\mathbb{E}\left(\bar{X}\right) = A \qquad \text{şi} \qquad \mathbb{E}\left(\frac{1}{n} \sum_{i=0}^{n-1} X_i^2 - \bar{X}^2\right) = \sigma^2 + A^2 - \mathbb{E}(\bar{X}^2) = \sigma^2 + A^2 - A^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

Multiplicând atunci statistica cu $n/(n-1)$, va fi nedeplasată pentru parametrul σ^2. Transformarea recomandată va fi:

$$\mathbf{g}(\mathbf{T}(V)) = \begin{pmatrix} \frac{1}{n}T_1(V) \\ \frac{1}{n-1}\left(T_2(V) - n\left(\frac{1}{n}T_1(V)\right)^2\right) \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \frac{1}{n-1}\left(\sum_{i=0}^{n-1} X_i^2 - n\bar{X}^2\right) \end{pmatrix}.$$

Un simplu calcul agebric arată că:

$$\sum_{i=0}^{n-1}(X_i - \bar{X})^2 = \sum_{i=0}^{n-1} X_i^2 - 2\sum_{i=0}^{n-1} X_i\bar{X} + n\bar{X}^2 = \sum_{i=0}^{n-1} X_i^2 - n\bar{X}^2,$$

iar estimatorul va fi:

$$\hat{\theta} = \begin{pmatrix} \bar{X} \\ \frac{1}{n-1}\sum_{i=0}^{n-1}(X_i - \bar{X})^2 \end{pmatrix},$$

care este un estimator MVU pentru $\theta = (A, \sigma^2)^t$. Factorul de normalizare $1/(n-1)$ pentru $\hat{\sigma}^2$ se datorează pierderii unui grad de libertate în estimarea mediei. În acest exemplu, $\hat{\theta}$ nu este un estimator eficient. Se poate demonstra (vom omite aici această demonstraţie) că $\bar{X}$ şi $\hat{\sigma}^2$ sunt statistici independente şi că

$$\bar{X} \sim \mathcal{N}\left(A, \frac{\sigma^2}{n}\right) \qquad şi \qquad \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1).$$

Matricea de covarianţă şi CRLB (vezi Exemplul 1.5) sunt:

$$C_{\hat{\theta}} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix} \qquad şi \qquad I^{-1}(\theta) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}.$$

Se observă că estimatorul MVU nu se putea obţine prin examinarea CRLB.

În final, observăm că dacă am fi ştiut sau am fi intuit că media de selecţie şi că dispersia de selecţie modificată sunt estimatori MVU, am fi putut simplifica studiul. Densitatea vectorului de selecţie este:

$$\begin{aligned} \mathcal{L}(v,\theta) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{n-1}(x_i - A)^2\right) \\ &= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{n-1}(x_i - \bar{x})^2 + n(\bar{x} - A)^2\right)\right)}_{=:\mathbf{g}(\mathbf{T}'(V),\theta)} \cdot \underbrace{1}_{=:h(v)}, \end{aligned}$$

unde $\mathbf{T}'(V) = (\bar{X}, \sum_{i=0}^{n-1}(X_i - \bar{X})^2)^t$. Împărţind a doua componentă la $(n-1)$ obţinem estimatorul $\hat{\theta}$. Desigur, există o transformare bijectivă între $\mathbf{T}'(V)$ şi $\mathbf{T}(V)$, ceea ce subliniază, încă o dată, că statisticile suficiente sunt unice până la o transformare bijectivă.

Pentru concluzia că $\hat{\theta}$ este un estimator MVU nu am verificat completitudinea statisticii. Aceasta rezultă datorită faptului că densitatea Gaussiană este un caz particular al familiilor de densităţi exponenţiale, despre care Kendall, Stuart [4], arată că sunt complete.

3 Estimatori liniari nedeplasaţi optimali (BLUE)

3.1 Introducere

În practică se întâmplă frecvent ca estimatorul MVU, dacă există, să nu poate fi găsit. De exemplu, este posibil să nu cunoaştem densitatea de repartiţie a datelor sau chiar să putem fi dispuşi să ne asumăm un model pentru acesta. În acest caz, metodele noastre anterioare, care se bazează pe CRLB şi teoria statisticilor suficiente, nu pot fi aplicate. Chiar dacă densitatea este cunoscută, ultimele abordări nu garantează obţinerea estimatorului

MVU. Confruntându-ne cu incapacitatea de a determina estimatorul *MVU* optimal, este rezonabil să apelăm la un estimator suboptimal. Procedând astfel, nu putem fi siguri de cât de multă performanţă am putea pierde (deoarece dispersia minimală a estimatorului *MVU* este necunoscută). Cu toate acestea, dacă dispersia estimatorului suboptimal poate fi stabilită şi dacă îndeplineşte specificaţiile sistemului nostru, atunci utilizarea sa poate fi justificată ca fiind adecvată pentru problema în cauză. Dacă dispersia sa este prea mare, atunci va trebui să ne uităm la alţi estimatori suboptimali, sperând să găsim unul care să îndeplinească specificaţiile noastre. O abordare frecventă este aceea de a restricţiona estimatorul să fie liniar în raport cu datele şi de a găsi estimatorul liniar care este nedeplasat şi are dispersia minimă. După cum vom vedea, acest estimator, care este denumit *estimator liniar nedeplasat optimal* (*BLUE, the best linear unbiased estimator*), poate fi determinat doar cu cunoaşterea momentelor de orinul întâi şi al doilea. Deoarece nu este necesară cunoaşterea completă a densităţii de repartiţie, *BLUE* este, de multe ori, cel mai potrivit pentru implementarea practică.

BLUE se bazează pe estimarea liniară definită de (56). Dacă impunem ca această estimare să fie nedeplasată, precum în (57), şi să minimizăm dispersia lui (58), atunci *BLUE* este dat de formula (60), iar dispersia minimală a acestui estimator este dată de aceeaşi formulă. Pentru aflarea estimatorului, este nevoie doar de cunoaşterea mediei şi dispersiei datelor. Atât în cazul scalar, cât şi în cel vectorial, dacă caracteristica este de tip Gaussian, atunci *BLUE* coincide cu *MVU.*

3.2 Modul de definire al BLUE

Considerăm o selecţie repetată de volum n asupra unei caracteristici a cărei densitate de repartiţie parametrizată este $f\left(x,\theta\right)$, vectorul empiric de selecţie fiind $v=\{x_1,x_2,...,x_n\}$. *BLUE* solicită restricţia ca estimatorul să fie de tip liniar. Mai precis, considerând $X_1,...,X_n$ variabilele de selecţie corespunzătoare eşantionării,

$$\hat{\theta}=\sum_{i=0}^{n-1}a_iX_i, \tag{56}$$

unde a_i sunt coeficienţi constanţi, ce trebuie determinaţi. În funcţie de alegerea acestora, putem genera diferiţi estimatori pentru θ. Cu toate acestea, cel mai bun estimator (*BLUE*) este definit ca fiind acela care este nedeplasat şi care are dispersia minimă. Deoarece ne limităm la clasa estimatorilor liniari, *BLUE* va fi optimal (cu alte cuvinte, este estimator *MVU*) doar atunci când *MVU* este, la rândul său, liniar. De exemplu, pentru problema estimării valorii mărimii curentului continuu în WGN (vezi Exemplul 1.3), estimatorul *MVU* este media de selecţie:

$$\hat{\theta}=\bar{X}=\sum_{i=0}^{n-1}\frac{1}{n}X_i,$$

care este, evident, de tip liniar. Prin urmare, dacă ne restricţionăm doar la clasa estimatorilor ce au această proprietate, nu vom pierde nimic deoarece estimatorul *MVU* aparţine acestei clase. Pe de altă parte, pentru problema estimării valorii medii a zgomotului repartizat uniform, estimatorul *MVU* este*:*

$$\hat{\theta}=\frac{n+1}{2n}\max_{i\in\{0,1,...,n-1\}}X_i,$$

care este neliniar în raport cu datele. Dacă am restricţiona căutarea doar la cazul estimatorilor liniari, atunci *BLUE* ar trebui să fie media de selecţie, aşa cum vom vedea în această secţiune. După cum este prezentat în Exemplul 2.8, diferenţa în performanţă este substanţială. Din nefericire, fără cunoaştere densităţii de repartiţie a statisticii, nu putem determina pierderea de performanţă prin apelarea la *BLUE*. Pentru anumite probleme de estimare, utilizarea lui *BLUE* poate fi total nepotrivită. Dacă privim la problema estimării puterii unui WGN, se poate arăta uşor că estimatorul *MVU* este (vezi Exemplul 1.5):

$$\widehat{\sigma^2}=\frac{1}{n}\sum_{i=0}^{n-1}X_i^2,$$

care este neliniar. Dacă am forţa ca estimatorul să fie liniar, similar formulei (56), $\widehat{\sigma^2}=\sum_{i=1}^{n}a_iX_i$, iar valoarea sa medie este:

$$\mathbb{E}(\widehat{\sigma^2})=\sum_{i=0}^{n-1}a_i\mathbb{E}\left(X_i\right)=0,\quad \text{deoarece}\quad \mathbb{E}\left(X_i\right)=0,\ \text{pentru fiecare } i\in\{1,2,...,n\}.$$

Nu putem deci găsi niciun estimator liniar care să fie nedeplasat, cu atât mai puţin unul de dispersie minimă. Deşi *BLUE* nu este potrivit acestei probleme, un *BLUE* ce utilizează datele modificate $Y_i=X_i^2$ va produce un

estimator viabil, deoarece pentru

$$\widehat{\sigma^2} = \sum_{i=0}^{n-1} a_i Y_i = \sum_{i=0}^{n-1} a_i X_i^2, \qquad \text{avem} \qquad \mathbb{E}(\widehat{\sigma^2}) = \sum_{i=0}^{n-1} a_i \sigma^2 = \sigma^2.$$

Există multe valori ale coeficienţilor a_i care ar satisface această restricţie. Problema este care valori trebuie utilizate pentru a obţine *BLUE*?

3.3 Determinarea lui BLUE

Pentur a determina *BLUE*, constrângem estimatorul $\hat{\theta}$ să fie liniar şi nedeplasat, iar apoi determinăm coeficienţii a_i care duc la minimizarea dispersiei. Avem, astfel,

$$\mathbb{E}(\hat{\theta}) = \sum_{i=0}^{n-1} a_i \mathbb{E}\left(X_i\right) = \theta. \tag{57}$$

Dispersia estimatorului $\hat{\theta}$ este

$$D^2(\hat{\theta}) = \mathbb{E}\left(\left(\sum_{i=0}^{n-1} a_i X_i - \mathbb{E}\left(\sum_{i=0}^{n-1} a_i X_i\right)\right)^2\right).$$

Folosind (57) şi utilizând abordarea vectorială $\mathbf{a} = \left(a_1, a_2, ..., a_n\right)^t$, obţinem:

$$D^2(\hat{\theta}) = \mathbb{E}\left(\left(\mathbf{a}^t\mathbf{X} - \mathbf{a}^t\mathbb{E}\left(\mathbf{X}\right)\right)^2\right) = \mathbb{E}\left(\left(\mathbf{a}^t\left(\mathbf{X} - \mathbb{E}\left(\mathbf{X}\right)\right)\right)^2\right) = \mathbb{E}\left(\mathbf{a}^t\left(\mathbf{X} - \mathbb{E}\left(\mathbf{X}\right)\right)\left(\mathbf{X} - \mathbb{E}\left(\mathbf{X}\right)\right)^t\mathbf{a}\right) = \mathbf{a}^t C \mathbf{a}. \tag{58}$$

Vectorul ponderilor $\mathbf{a}$ se obţine prin minimzarea (58), în raport cu restricţia (57). Înainte de a continua, trebuie să presupunem o anumită formă a mediei $\mathbb{E}\left(X_i\right)$. Pentru a satisface condiţia de nedeplasare, $\mathbb{E}\left(X_i\right)$ ar trebui să fie liniar în raport cu θ, adică

$$\mathbb{E}\left(X_i\right) = s_i\theta, \quad i \in \{0, 1, ..., n-1\}, \tag{59}$$

unde coeficienţii s_i sunt cunoscuţi. Altfel, ar putea fi imposibil ca restricţia să fie satisfăcută. De exemplu, dacă $\mathbb{E}\left(X_i\right) = \cos\theta$, atunci condiţia de nedeplasare devine:

$$\sum_{i=0}^{n-1} a_i \cos\theta = \theta, \quad \text{pentru toţi } \theta.$$

Pentru acest exemplu este clar că nu există coeficienţii a_i care să satisfacă egalitatea anterioară. Observăm că, dacă scriem pe X_i sub forma $X_i = \mathbb{E}\left(X_i\right) + \left(X_i - \mathbb{E}\left(X_i\right)\right)$, atunci, notând $W_i := X_i - \mathbb{E}\left(X_i\right)$ ca fiind zgomotul pe sistem,

$$X_i = \theta s_i + W_i, \quad \text{pentru toţi } i \in \{1, 2, ..., n\}.$$

Pentru a generaliza analiza, avem nevoie de o transformare neliniară a datelor observate, aşa cum am descris deja. Cu ipoteza dată de (59), problema estimării se pune în modul următor. Pentru a determina *BLUE*, trebuie să minimizăm dispersia $D^2(\hat{\theta}) = \mathbf{a}^t C \mathbf{a}$, cu restricţia de nedeplasare, care, datorită formulelor (57) şi (59) devine:

$$\sum_{i=0}^{n-1} a_i \mathbb{E}\left(X_i\right) = \theta \iff \sum_{i=0}^{n-1} a_i s_i \theta = \theta \iff \sum_{i=0}^{n-1} a_i s_i = 1 \iff \mathbf{a}^t\mathbf{s} = 1,$$

unde $\mathbf{s} := \left(s_0, s_1, ..., s_{n-1}\right)^t$. Vom vedea că soluţia acestei probleme de minimizare este:

$$\mathbf{a}_{opt} = \frac{C^{-1}\mathbf{s}}{\mathbf{s}^t C^{-1}\mathbf{s}},$$

iar *BLUE* are reprezentarea

$$\hat{\theta} = \frac{\mathbf{s}^t C^{-1}\mathbf{X}}{\mathbf{s}^t C^{-1}\mathbf{s}}, \qquad \text{dispersia sa fiind dată de} \qquad D^2(\hat{\theta}) = \frac{1}{\mathbf{s}^t C^{-1}\mathbf{s}}. \tag{60}$$

În conformitate cu (59), deoarece $\mathbb{E}\left(\mathbf{X}\right) = \theta\mathbf{s}$, *BLUE* este nedeplasat:

$$\mathbb{E}(\hat{\theta}) = \frac{\mathbf{s}^t C^{-1}\mathbb{E}(\mathbf{X})}{\mathbf{s}^t C^{-1}\mathbf{s}} = \frac{\mathbf{s}^t C^{-1}\theta\mathbf{s}}{\mathbf{s}^t C^{-1}\mathbf{s}} = \theta.$$

Pentru a determina *BLUE*, trebuie să cunoaştem doar pe $\mathbf{s}$ şi C (matricea de covarianţă) sau primele două momente, dar nu şi întreaga densitate de repartiţie, după cum vom vedea în exemplele prezentate în continuare.

Pentru moment, prezentăm modul de obţinere a valorii $\mathbf{a}_{opt}$, rezultată în urma rezolvării unei probleme de optimizare:

$$\begin{cases} \min D^2(\hat{\theta}) = \min \mathbf{a}^t C \mathbf{a} \\ \mathbf{a}^t \mathbf{s} = \mathbf{1}. \end{cases}$$

Vom folosi metoda multiplicatorilor lui Lagrange. Funcţia Lagrangiană este $J = \mathbf{a}^t C \mathbf{a} + \lambda \left(\mathbf{a}^t \mathbf{s} - \mathbf{1} \right)$, al cărei gradient, în raport cu vectorul $\mathbf{a}$, este $\partial J / \partial \mathbf{a} = 2C\mathbf{a} + \lambda \mathbf{s}$. Egalându-l cu zero, obţinem:

$$\mathbf{a} = -\frac{\lambda}{2} C^{-1} \mathbf{s}, \qquad \text{iar} \qquad \mathbf{a}^t \mathbf{s} = -\frac{\lambda}{2} \mathbf{s}^t C^{-1} \mathbf{s} = \mathbf{1} \quad \Longrightarrow \quad -\frac{\lambda}{2} = \frac{1}{\mathbf{s}^t C^{-1} \mathbf{s}}.$$

Rezultă deci că

$$\mathbf{a}_{opt} = \frac{C^{-1}\mathbf{s}}{\mathbf{s}^t C^{-1} \mathbf{s}} \quad \text{şi} \quad D^2(\hat{\theta}) = \mathbf{a}_{opt}^t C \mathbf{a}_{opt} = \frac{\mathbf{s}^t C^{-1} C C^{-1} \mathbf{s}}{\left(\mathbf{s}^t C^{-1} \mathbf{s} \right)^2} = \frac{1}{\mathbf{s}^t C^{-1} \mathbf{s}}.$$

Pentru a vedea faptul că $\mathbf{a}_{opt}$ este minimul global şi nu doar un minim local considerăm funcţia:

$$\begin{aligned} G &= \left(\mathbf{a} - \mathbf{a}_{opt} \right)^t C \left(\mathbf{a} - \mathbf{a}_{opt} \right) = \mathbf{a}^t C \mathbf{a} - 2 \mathbf{a}_{opt}^t C \mathbf{a} + \mathbf{a}_{opt}^t C \mathbf{a}_{opt} \\ &= \mathbf{a}^t C \mathbf{a} - 2 \frac{\mathbf{s}^t C^{-1} C \mathbf{a}}{\mathbf{s}^t C^{-1} \mathbf{s}} + \frac{1}{\mathbf{s}^t C^{-1} \mathbf{s}} = \mathbf{a}^t C \mathbf{a} - \frac{1}{\mathbf{s}^t C^{-1} \mathbf{s}}. \end{aligned}$$

Prin urmare,

$$\mathbf{a}^t C \mathbf{a} = \left(\mathbf{a} - \mathbf{a}_{opt} \right)^t C \left(\mathbf{a} - \mathbf{a}_{opt} \right) + \frac{1}{\mathbf{s}^t C^{-1} \mathbf{s}},$$

care este minimizată doar de $\mathbf{a} = \mathbf{a}_{opt}$. Aceasta se întâmplă deoarece proprietatea de pozitivă definire a matricei C face ca primul termen din membrul din dreapta al egalităţii anterioare să fie strict mai mare decât zero pentru $\mathbf{a} - \mathbf{a}_{opt} \neq \mathbf{0}$ (vectorul nul).

Exemplul 3.1 ***Mărimea curentului continuu în WGN.***

(a) Presupunem că datele observate sunt date de $X_i = A + W_i$, $i \in \{0, 1, ..., n-1\}$, unde, pentru fiecare i, W_i este un zgomot alb, având dispersia σ^2 (şi densitatea de repartiţie necunoscută). Problema constă în estimarea mărimii A a curentului. Cum W_i nu sunt, neapărat, de tip Gaussian, zgomotele pot fi statistic dependente, chiar dacă ele sunt, două câte două, necorelate. Deoarece $\mathbb{E}(X_i) = A$, atunci, conform relaţiei (59), $s_i = 1$ şi, prin urmare, $\mathbf{s} = \mathbf{1}$. Reprezentarea (60) furnizează forma lui BLUE:

$$\hat{A} = \frac{\mathbf{1}^t \frac{1}{\sigma^2} I \, \mathbf{X}}{\mathbf{1}^t \frac{1}{\sigma^2} I \, \mathbf{1}} = \frac{1}{n} \sum_{i=0}^{n-1} X_i = \bar{X}, \qquad \textit{având dispersia} \qquad D^2(\hat{A}) = \frac{1}{\mathbf{1}^t \frac{1}{\sigma^2} \mathbf{1}} = \frac{\sigma^2}{n}.$$

Prin urmare, media de selecţie este BLUE, afirmaţie ce este adevărată independent de densitatea de repartiţie a caracteristicii. În plus, este estimatorul de dispersie minimă pentru zgomotul Gaussian.

(b) Presupunem acum că toate W_i, sunt zgomote albe necorelate, de medie zero şi cu dispersia $D^2(W_i) = \sigma_i^2$, $i \in \{0, 1, ..., n-1\}$. La fel ca şi la punctul precedent, $\mathbf{s} = \mathbf{1}$ şi, conform cu (60),

$$\hat{A} = \frac{\mathbf{1}^t C^{-1} \mathbf{X}}{\mathbf{1}^t C^{-1} \mathbf{1}}, \qquad \textit{cu dispersia} \qquad D^2(\hat{\theta}) = \frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}}.$$

Matricea de covarianţă, respectiv inversa sa, sunt date de formulele:

$$C = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \qquad \textit{şi} \qquad C^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{pmatrix}.$$

Prin urmare,

$$\hat{A} = \frac{\sum_{i=0}^{n-1} \frac{X_i}{\sigma_i^2}}{\sum_{i=0}^{n-1} \frac{1}{\sigma_i^2}} \qquad \textit{şi} \qquad D^2(\hat{A}) = \frac{1}{\sum_{i=0}^{n-1} \frac{1}{\sigma_i^2}}. \tag{61}$$

În concluzie, BLUE ponderează mai mult acele probe ce au dispersia mai mică, în incercarea de a egaliza contribuţia zgomotelor provenite de la fiecare probă în parte. Numitorul din formulele (61) este factorul de ajustare ce face ca estimatorul obţinut să fie nedeplasat.

Datorită ipotezei zgomotului Gaussian, estimatorul poate fi considerat a fi eficient şi, prin urmare este un estimator *MVU*. Nu este doar o coincidenţă faptul că estimatorul *MVU* pentru cazul zgomotului Gasussian şi *BLUE* sunt aceeaşi. Avem de a face, de fapt, cu un rezultat general, care afirmă că, în estimarea parametrilor unui model liniar, cele două tipuri de estmatori coincid dacă ne situăm în ipoteza menţionată privind tipul zgomotului pe sistem.

3.4 Cazul parametrului multidimensional

Dacă parametrul estimat este un vector $p \times 1$ dimensional, atunci, pentru ca estimatorul să fie liniar în raport cu datele ar trebui ca el să aiba forma:

$$\hat{\theta}_i = \sum_{i=0}^{n-1} a_{in} X_i, \quad i \in \{1, 2, ..., p\}, \tag{62}$$

unde a_{in} sunt coeficienţii de ponderare. Scrisă sub formă matriceală, formula (62) devine $\hat{\theta} = \mathbf{A}V$, unde $\mathbf{A} \in \mathcal{M}_{p\times n}(\mathbb{R})$. Pentru ca estimatorul $\hat{\theta}$ să fie nedeplasat, trebuie ca

$$\mathbb{E}(\hat{\theta}_i) = \sum_{i=0}^{n-1} a_{in}\mathbb{E}(X_i) = \theta_i, \quad i \in \{1, 2, ..., p\}, \quad \text{sau, echivalent,} \quad \mathbb{E}(\hat{\theta}) = \mathbf{A}\mathbb{E}(V) = \theta. \tag{63}$$

Un estimator liniar este potrivit doar pentru probleme în care restricţia privind nedeplasarea poate fi satisfăcută. Din formula matriceală precedentă obţinem că:

$$\mathbb{E}(V) = \mathbf{H}\theta, \quad \text{unde } \mathbf{H} \in \mathcal{M}_{n\times p}(\mathbb{R}). \tag{64}$$

În cazul parametrului scalar (vezi (59)), $\mathbb{E}(V) = \mathbf{H}\theta = (s_0, s_1, ..., s_{n-1})^t\theta$. Substituind (64) în (63), obţinem condiţia de nedeplasare

$$\mathbf{AH} = I. \tag{65}$$

Dacă definim linia $\mathbf{a}_i := \left(a_{i0}, a_{i1}, ..., a_{i(n-1)}\right)^t$, astfel încât $\hat{\theta}_i = \mathbf{a}_i^t V$, atunci condiţia de nedeplasare poate fi rescrisă pentru fiecare $\mathbf{a}_i$ notând

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_p^t \end{pmatrix} \quad \text{şi considerând } \mathbf{h}_i \text{ coloana } i \text{ a lui } \mathbf{H} = \begin{pmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_p \end{pmatrix}.$$

Astfel, condiţia de nedeplasare (65) se reduce la

$$\mathbf{a}_i^t\mathbf{h}_j = \delta_{ij}, \quad i, j \in \{1, 2, ..., p\}. \tag{66}$$

Dispersia estimatorilor va fi, pentru fiecare i,

$$D^2(\hat{\theta}_i) = \mathbf{a}_i^t C \mathbf{a}_i. \tag{67}$$

BLUE pentru parametrul vectorial se obţine prin minimizarea cantităţilor (67), cu restricţia dată de (66), pentru fiecare i. Arătăm în continuare că *BLUE* şi matricea de covarianţă sunt date de formulele:

$$\hat{\theta} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}\mathbf{H}^t C^{-1} V, \qquad \text{iar} \qquad C_{\hat{\theta}} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}. \tag{68}$$

Ne situăm în ipoteza unui model liniar, $V = \mathbf{H}\theta + \mathbf{w}$, cu $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, C)$. Scopul este deci acela de a estima parametrul vectorial θ, unde $\mathbb{E}(V) = \mathbf{H}\theta$. Aceasta este exact presupunerea dată de formula (64). Dar, estimatorul *MVU* pentru date de tip Gaussian este

$$\hat{\theta} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}\mathbf{H}^t C^{-1} V,$$

care este liniar în V. Putem concluziona că dacă datele sunt de tip Gaussian, atunci *BLUE* este şi estimator *MVU*.

Problema minimizării dispersiilor date de (67), în prezenţa restricţiilor de nedeplasare (66) este similară cazului scalar, exceptând prezenţa restricţiilor asupra vectorilor $\mathbf{a}_i$. Avem acum p restricţii asupra acestor vectori, nu doar una. Cum vectorii $\mathbf{a}_i$ pot lua orice valori, independent între ei, avem de a face, de fapt, cu p probleme de minimizare separate, interconectate doar de restricţii. Considerăm, pentru fiecare i, Lagrangian-ul

$$J_i = \mathbf{a}_i^t C \mathbf{a}_i + \sum_{j=1}^{p} \lambda_j^{(i)} (\mathbf{a}_i^t \mathbf{h}_j - \delta_{ij}), \quad \text{cu derivatele parţiale} \quad \frac{\partial J_i}{\partial \mathbf{a}_i} = 2C\mathbf{a}_i + \sum_{j=1}^{p} \lambda_j^{(i)} \mathbf{h}_j = 2C\mathbf{a}_i + \mathbf{H}\lambda_i,$$

pentru ultima egalitate folosind notaţiile $\lambda_i := (\lambda_1^{(i)}, \lambda_2^{(i)}, ..., \lambda_p^{(i)})^t$ şi $\mathbf{H} = (\ \mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_p\ )$. Egalând gradientul cu zero, găsim:

$$\mathbf{a}_i = -\frac{1}{2} C^{-1} \mathbf{H} \lambda_i. \tag{69}$$

Pentru a determina multiplicatorii lui Lagrange, ne vom folosi de restricţiile (66) care, în formă combinată, duc la ecuaţia:

$$\begin{pmatrix} \mathbf{h}_1^t \\ \vdots \\ \mathbf{h}_{i-1}^t \\ \mathbf{h}_i^t \\ \mathbf{h}_{i+1}^t \\ \vdots \\ \mathbf{h}_p^t \end{pmatrix} \mathbf{a}_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Notăm cu e_i vectorul canonic i şi obţinem restricţia $\mathbf{H}^t \mathbf{a}_i = e_i$. Folosim (69), iar restricţia devine:

$$\mathbf{H}^t \mathbf{a}_i = -\frac{1}{2} \mathbf{H}^t C^{-1} \mathbf{H} \lambda_i = e_i.$$

Dacă presupunem inversabilitatea matricei $\mathbf{H}^t C^{-1} \mathbf{H}$, atunci vectorul multiplicatorilor lui Lagrange este:

$$-\frac{1}{2}\lambda_i = (\mathbf{H}^t C^{-1} \mathbf{H})^{-1} e_i, \qquad \text{iar} \qquad \mathbf{a}_{i_{opt}} = C^{-1}\mathbf{H}(\mathbf{H}^t C^{-1}\mathbf{H})^{-1} e_i. \tag{70}$$

Dispersia acestui estimator este:

$$D^2(\hat{\theta}_i) = \mathbf{a}_{i_{opt}}^t C \mathbf{a}_{i_{opt}} = e_i^t (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} C C^{-1} \mathbf{H} (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} e_i = e_i^t (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} e_i.$$

Similar cazului scalar, se poate arăta că $\mathbf{a}_{i_{opt}}$ este un punct de minim global. Putem exprima vectorul *BLUE* sub o formă mai compactă, astfel:

$$\hat{\theta} = \begin{pmatrix} \mathbf{a}_{1_{opt}}^t V \\ \mathbf{a}_{2_{opt}}^t V \\ \vdots \\ \mathbf{a}_{p_{opt}}^t V \end{pmatrix} = \begin{pmatrix} e_1^t (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} V \\ e_2^t (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} V \\ \vdots \\ e_p^t (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} V \end{pmatrix} = \begin{pmatrix} e_1^t \\ e_2^t \\ \vdots \\ e_p^t \end{pmatrix} (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} V = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} V,$$

deoarece $(e_1^t\ e_2^t, ..., e_p^t)$ este matricea unitate. De asemenea, matricea de covarianţă a estimatorului vectorial $\hat{\theta}$ este:

$$C_{\hat{\theta}} = \mathbb{E}\left((\hat{\theta} - \mathbb{E}(\hat{\theta}))(\hat{\theta} - \mathbb{E}(\hat{\theta}))^t \right),$$

unde

$$\hat{\theta} - \mathbb{E}(\hat{\theta}) = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} (\mathbf{H}\theta + \mathbf{w}) - \mathbb{E}(\hat{\theta}) = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} \mathbf{w}.$$

Obţinem astfel:

$$\begin{aligned} C_{\hat{\theta}} &= \mathbb{E}\left((\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} \mathbf{w}\mathbf{w}^t C^{-1} \mathbf{H} (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \right) \\ &= (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} \mathbf{H}^t C^{-1} C C^{-1} \mathbf{H} (\mathbf{H}^t C^{-1}\mathbf{H})^{-1} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}, \end{aligned}$$

deoarece matricea de covarianţă a lui $\mathbf{w}$ este C. Minimul dispersiilor este dat de către elementele diagonale ale matricei $C_{\hat{\theta}}$, adică:

$$D^2(\hat{\theta}_i) = e_i^t C_{\hat{\theta}} e_i = (\mathbf{H}^t C^{-1}\mathbf{H})_{ii}^{-1}, \quad i \in \{1, 2, ..., p\},$$

obţinând astfel rezultatul dorit.

În cadrul considerat, modelul liniar general nu presupune existenţa zgomotului Gaussian. Datele au doar o formă liniară de reprezentare. Următorul rezultat oferă formula estimatorului şi a matricei sale de covarianţă în această situaţie.

Teorema 6 (Gauss-Markov) *Dacă datele au forma liniară generală:*

$$V = \mathbf{H}\theta + \mathbf{w}, \quad \textit{unde } \mathbf{H} \in \mathcal{M}_{n\times p}(\mathbb{R}) \textit{ este cunoscută}, \quad \theta \in \mathcal{M}_{p\times 1} \textit{ este vectorul parametrilor,}$$

iar $\mathbf{w}$ *este vectorul zgomotelor pe sistem, de tip* $n \times 1$*, având media zero şi matricea de covarianţă* C *(densitatea de repartiţie fiind necunoscută, nu neapărat de tip Gaussian), atunci BLUE pentru* θ *este:*

$$\hat{\theta} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}\mathbf{H}^t C^{-1} V, \quad \textit{iar dispersia minimă pentru } \hat{\theta}_i \textit{ este } D^2(\hat{\theta}_i) = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}_{ii},\ i \in \{1, 2, ..., p\}. \tag{71}$$

În plus, matricea de covarianţă a estimatorului vectorial $\hat{\theta}$ *este:*

$$C_{\hat{\theta}} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}.$$

3.5 Un exemplu privind procesarea semnalului. Localizarea sursei

O problemă de interes privind traficul aerian îl constituie determinarea poziţiei unei surse emitente a unui semnal. De exemplu, pentru a determina poziţia unui avion, se utilizează, de regulă, o reţea de antene. Metoda constă în estimarea diferenţelor temporale dintre o serie de măsurători (*TDOA, time difference of arrival*). Astfel, semnalul emis de un avion (sursa) este recepţionat, în acelaşi timp, de trei antene din reţea, având deci $t_1 = t_2 = t_3$. *TDOA* $t_2 - t_1$ dintre antena 1 şi 2 este zero, la fel ca şi $t_3 - t_2$ dintre antenele 2 şi 3. Putem afirma astfel că range-urile (razele de acţiune, distanţele) R_i către cele trei antene sunt aceleaşi. Intersecţia porţiunilor de disc având ca centru antenele şi ca rază range-urile R_i, reprezintă posibila locaţie a avionului emitent al semnalului.

Vom examina problema de localizare folosind teoria estimaţiei. Considerăm o reţea de n antene ce sunt poziţionate în locaţii cunoscute şi că măsurătorile timpilor de sosire (recepţionare a semnalului de către antene) t_i, $i \in \{0, 1, ..., n-1\}$ sunt disponibile. Trebuie să estimăm poziţia sursei (x_s, y_s). De asemenea, presupunem că timpii de sosire sunt perturbaţi de câte un zgomot aleator de medie zero, dispersie cunoscută, dar densitate de repartiţie necunoscută. Pentru semnalul emis de sursă la momentul $t = T_0$, măsurătorile sunt modelate de relaţiile:

$$t_i = T_0 + R_i/c + \varepsilon_i, \quad i \in \{0, 1, ..., n-1\}, \tag{72}$$

unde ε_i sunt zgomotele (perturbaţiile) măsurătorilor, iar c reprezintă viteza de propagare. Zgomotele măsurătorilor sunt de medie 0, dispersie σ^2 şi necorelate între ele. Presupunem că *nu există nicio informaţie privind distribuţia zgomotelor*. Vom stabili o relaţie între raza de acţiune R_i a fiecărei antene (distanţa până la ţintă) şi poziţia necunoscută $\theta = \begin{pmatrix} x_s & y_s \end{pmatrix}^t$ a emiţătorului. Notăm poziţia fiecărei antene cu $(x_i, y_i)_i$, poziţii pe care le presupunem cunoscute şi avem:

$$R_i = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} \tag{73}$$

Dacă introducem formula lui R_i din (73) în (72), se observă că modelul este neliniar în raport cu parametrii necunoscuţi x_s şi y_s. Pentru a aplica Teorema Gauss-Markov, presupunem că este disponibilă o poziţie probabilă a sursei, (x_m, y_m). Această poziţie probabilă, care este apropiată de adevărata poziţie a sursei (poziţia exactă) poate fi determinată din măsurători anterioare. O astfel de situaţie apare, de regulă, când sursa este urmărită de receptori. În consecinţă, pentru a estima poziţia nouă a sursei, trebuie să estimăm parametrul $\theta = \begin{pmatrix} x_s - x_m & y_s - y_m \end{pmatrix}^t = \begin{pmatrix} \delta x_s & \delta y_s \end{pmatrix}^t$. Fie R_{m_i} razele de acţiune nominale, precum în Figura 5.

Folosim o dezvoltare Taylor de ordinul întâi a cantităţii R_i (considerată ca o funcţie de două variabile, x_s şi y_s), în jurul poziţiei nominale $x_s = x_m, y_s = y_m$:

$$R_i \approx R_{m_i} + \frac{x_m - x_i}{R_{m_i}}\delta x_s + \frac{y_m - y_i}{R_{m_i}}\delta y_s. \tag{74}$$

Folosim această aproximare în formula (72) şi obţinem:

$$t_i = T_0 + \frac{R_{m_i}}{c} + \frac{x_m - x_i}{R_{m_i} c}\delta x_s + \frac{y_m - y_i}{R_{m_i} c}\delta y_s + \varepsilon_i, \quad i \in \{0, 1, ..., n-1\},$$

formulă care este, acum, liniară în raport cu parametrii δx_s şi δy_s.

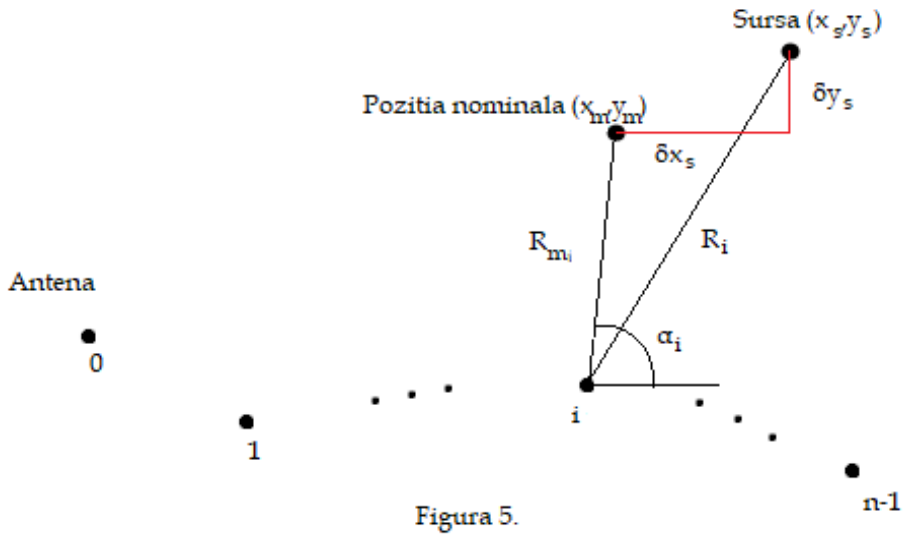


Figura 5.

De asemenea, deoarece

$$\cos\alpha_i = \frac{x_m - x_i}{R_{m_i}} \qquad \text{şi} \qquad \sin\alpha_i = \frac{y_m - y_i}{R_{m_i}},$$

modelul se poate scrie, simplificat,

$$t_i = T_0 + \frac{R_{m_i}}{c} + \frac{\cos\alpha_i}{c}\delta x_s + \frac{\sin\alpha_i}{c}\delta y_s + \varepsilon_i, \quad i \in \{0, 1, ..., n-1\}.$$

Termenul R_{m_i}/c este o constantă cunoscută care poate fi încorporată în măsurătoare prin notaţia

$$\tau_i = t_i - \frac{R_{m_i}}{c}.$$

Modelul nostru linear devine:

$$\tau_i = T_0 + \frac{\cos\alpha_i}{c}\delta x_s + \frac{\sin\alpha_i}{c}\delta y_s + \varepsilon_i, \quad i \in \{0, 1, ..., n-1\}, \tag{75}$$

unde parametrii necunoscuţi sunt $T_0, \delta x_s, \delta y_s$. Presupunerea că momentul T_0 la care sursa emite un semnal este necunoscut este des întâlnită în cazurile concrete. Cunoaşterea lui T_0 ar presupune o sincronizare precisă a ceasurilor sursei şi receiver-ului, fapt care, în situaţii reale este de evitat. De regulă, se consideră *TDOA* care elimină pe T_0 din (75). Construim această diferenţă de timp *TDOA* în felul următor:

$$\begin{cases} \xi_1 &= \tau_1 - \tau_0 \\ \xi_2 &= \tau_2 - \tau_1 \\ &= \\ \xi_{n-1} &= \tau_{n-1} - \tau_{n-1}. \end{cases}$$

Formula (75) conduce la următorul model (final) liniar:

$$\xi_i = \frac{1}{c}\left(\cos\alpha_i - \cos\alpha_{i-1}\right)\delta x_s + \frac{1}{c}\left(\sin\alpha_i - \sin\alpha_{i-1}\right)\delta y_s + \varepsilon_i - \varepsilon_{i-1}, \quad i \in \{1, 2, ..., n-1\}. \tag{76}$$

Am redus problema estimării la cea descrisă de Teorema Gauss-Markov, unde

$$\theta = \begin{pmatrix} \delta x_s \\ \delta y_s \end{pmatrix}, \quad \mathbf{H} = \frac{1}{c}\begin{pmatrix} \cos\alpha_1 - \cos\alpha_0 & \sin\alpha_1 - \sin\alpha_0 \\ \cos\alpha_2 - \cos\alpha_1 & \sin\alpha_2 - \sin\alpha_1 \\ \vdots & \vdots \\ \cos\alpha_{n-1} - \cos\alpha_{n-2} & \sin\alpha_{n-1} - \sin\alpha_{n-2} \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} \varepsilon_1 - \varepsilon_0 \\ \varepsilon_2 - \varepsilon_1 \\ \vdots \\ \varepsilon_{n-1} - \varepsilon_{n-2} \end{pmatrix}.$$

Vectorul zgomotelor $\mathbf{w}$ are media zero, dar nu mai este compus din variabile aleatoare necorelate între ele. Pentru a determina matricea de covarianţă, observăm că:

$$\mathbf{w} = \mathbf{A}\varepsilon = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{n-1} \end{pmatrix}, \quad \mathbf{A} \in \mathcal{M}_{(n-1)\times n}(\mathbb{R}).$$

Deoarece matricea de covarianţă a lui ε este $\sigma^2 I$, obţinem:

$$C = \mathbb{E}(\mathbf{A}\varepsilon\varepsilon^t\mathbf{A}^t) = \sigma^2\mathbf{A}\mathbf{A}^t.$$

Din (71), *BLUE* pentru parametrul vectorial al poziţiei sursei este:

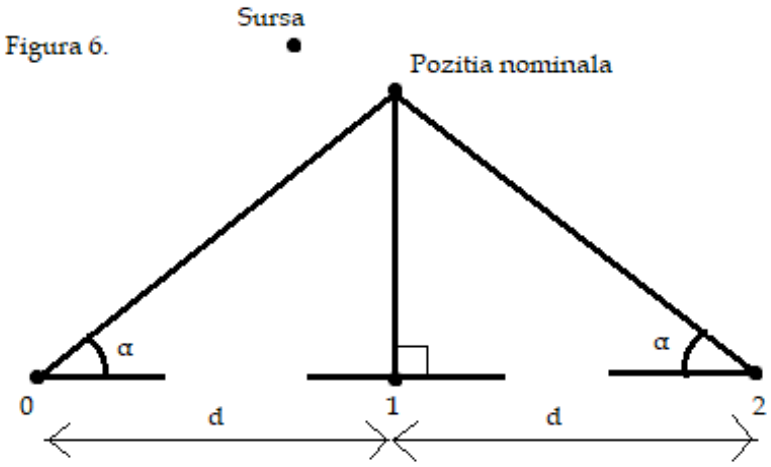
$$\hat{\theta} = (\mathbf{H}^t C^{-1}\mathbf{H})^{-1}\mathbf{H}^t C^{-1}\xi = (\mathbf{H}^t(\mathbf{A}\mathbf{A}^t)\mathbf{H})^{-1}\mathbf{H}^t(\mathbf{A}\mathbf{A}^t)^{-1}\xi, \tag{77}$$

iar dispersia minimă, respectiv matricea de covarianţă, sunt date de:

$$D^2(\hat{\theta}_i) = \sigma^2\left(\left(\mathbf{H}^t(\mathbf{A}\mathbf{A}^t)^{-1}\mathbf{H}\right)^{-1}\right)_{ii} \qquad \text{şi} \qquad C_{\hat{\theta}} = \sigma^2\left(\mathbf{H}^t(\mathbf{A}\mathbf{A}^t)^{-1}\mathbf{H}\right)^{-1}. \tag{78}$$

Ca un exemplu, în Figura 6, avem o reţeaua de 3 antene în linie, care este, de fapt, numărul minim de antene necesare pentru un model de localizare.

Figura 6.



Pentru acest model avem:

$$\mathbf{H} = \frac{1}{c}\begin{pmatrix} -\cos\alpha & 1-\sin\alpha \\ -\cos\alpha & -(1-\sin\alpha) \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

iar matricea de covarianţă este, conform formulei (78),

$$C_{\hat{\theta}} = \sigma^2 c^2 \begin{pmatrix} \dfrac{1}{2\cos^2\alpha} & 0 \\ 0 & \dfrac{3/2}{(1-\sin\alpha)^2} \end{pmatrix}.$$

Pentru o mai bună localizare, unghiul α ar tebui să fie cât mai mic. Aceasta se poate realiza mărind spaţiul d dintre antene. Remarcăm, de asemenea, că acurateţea locaţiei este dependentă de range, acurateţea localizării îmbunătăţindu-se pentru range-uri scurte sau pentru unghiuri α mici.

Bibliografie

[1] Devore, J; Berk, K., *Modern Mathematical Statistics with Applications*, 2^{nd} Edition, Springer New York Dordrecht Heidelberg London, 2012.

[2] Kay, S., *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall; 1 edition, April 5, 1993.

[3] Kay, S., *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, Pearson; 1 edition, February 6, 1998.

[4] Kendall, M.G.; Stuart, A., *The Advanced Theory of Statistics*, Volume 2, Inference and Relationships, Hafner Publishing Company, 1961 (Edition by Wiley, 2010).

[5] Montgomery, D; Runger, G, *Applied Statistics and Probability for Engineers*, 3^{rd} Edition, John Wiley & Sons, Inc, 2003.

[6] Wackerly, D.; Mendenhall, W.; Scheaffer, R., *Mathematical Statistics with Applications*, 7^{th} Edition, Thomson Brooks/Cole, 2008.